Supported diagnosis of ADHD from EEG signals based on Hidden Markov Models and Probability Product Kernels

A Thesis Presented for the Master of Science in Electrical Engineering Degree

Student: María Camila Maya Piedrahita
Director: Álvaro Ángel Orozco Gutiérrez, PhD
Co-director: David Augusto Cárdenas Peña, PhD

Universidad Tecnológica de Pereira
Engineering Faculty - Electrical Engineering Program
Master in Electrical Engineering
Research Group in Automática
Pereira, Risaralda, Colombia
2021

Abstract

Attention deficit hyperactivity disorder (ADHD), most often present in childhood, may persist in adult life, hampering personal development. However, ADHD diagnosis is a real challenge since it highly depends on the clinical observation of the patient, the parental and scholar information, and the specialist expertise. Despite demanded objective diagnosis aids from biosignals, the physiological biomarkers lack robustness and significance under the non-stationary and non-linear electroencephalographic dynamics. Therefore, this work presents a supported diagnosis methodology for ADHD from the dynamic characterization of EEG based on hidden Markov models (HMM) and probability product kernels (PPK). Based on the symptom of impulsivity, the proposed approach trains an HMM for each subject from EEG signals in failed inhibition tasks. In the first instance, PPK measures the similarity between subjects through the inner product between their trained HMMs. Then, given the computational costs, fast computation of PPK for HMM facilitates parameter tuning of kernel similarity. Finally, the Kernel Principal Component Analysis (KPCA) projects the PPK to a lower-dimensional space, allowing the interpretability of the results. Thus, a support vector machine supports the diagnosis of ADHD as a classification task using PPK as the inner product operator. The methodology compared classification results on EEG signals with all channels, channels of interest (COI), and analysis in the *Theta*, *Alpha*, and *Beta* frequency bands. The results show an accuracy rate of 97.0% in the *Beta* band in COI, which supports the assumption that this frequency rhythm may be correlated to differences between ADHD and controls regarding attentional allocation during the execution of the cognitive task.

Aknowledgments

To my director Álvaro Ángel Orozco Gutiérrez and my professor David Augusto Cárdenas Peña for the time and dedication they gave me to achieve the objectives of this research.
To my parents José Fernando Maya and Luz Dary Piedrahita for their dedication, and love.
To my uncle Ruben Dario Maya for his constant support.
This work was fully funded by MinCiencias projects: ”Herramienta de apoyo al diagnóstico del TDAH en niños a partir de múltiples características de actividad eléctrica cerebral desde registros EEG” with code number 111080763051. Authors also thank to the Vice-rectory for research from Universidad Tecnológica de Pereira for supporting the development of this work.

Contents

List of Figures

List of Tables

Chapter 1

List of Symbols and Abbreviations

1.1 Symbols

Symbol	Definition	Symbol	Definition
$\mathcal{X}$	Observation matrix	$\boldsymbol{X}_i$	Trial of each subject
N	Set time series	M	Number of state
$\boldsymbol{q}_i$	Sequence of hidden states	$\boldsymbol{\pi}$	Priors vector
$\boldsymbol{A}$	Transition matrix	$\boldsymbol{\mu}_m$	Means vector of a state
$\boldsymbol{\Sigma}_m$	Covars vector of a state	θ_i	Parameters HMM of a subject
T	Length of sequence	ρ	Power of PPK
$\kappa(\boldsymbol{X}_i, \boldsymbol{X}_j)$	Probability Product Kernel	$\boldsymbol{K}_{\mathcal{X}}$	PPK at all HMMs pairs
ψ_{mn}	Inner product between pdfs	$\boldsymbol{\alpha}_0$	Product between $\boldsymbol{\pi}$ and ψ
$\boldsymbol{\alpha}$	Product between $\boldsymbol{A}$ and ψ	$\kappa\left(\theta_i, \theta_j\right)$	Fast computation of PPK
$\boldsymbol{K}_\theta$	Similarity kernel by subjects	$\boldsymbol{K}_\psi$	Similarity kernel by states
$\phi(x)$	Nonlinear transformation	$\mathbf{C}$	Covariance matrix in feature space
λ	Eigenvalues	$\mathbf{V}$	Eigenvectors
$\mathbf{a}$	Coefficient vector	$\tilde{\mathbf{K}}$	Gramm matrix

1.2 Abbrevations

Abb	Definition	Abb	Definition
HMM	Hidden Markov Model	ADHD	Attention Deficit and Hyperactivity
EEG	Electroencephalography	RSST	Reward Stop-Signal Task
CIO	Channels of interest	ALL	All channels
BD	Behavior data	HC	Healthy control
IC	Increasing condition	DC	Decreasing condition
LL	Log-likelihood	PPK	Probability Product Kernel
SVM	Support Vector Machine	KPCA	Kernel Principal Components Analysis

Chapter 2

Introduction

2.1 Problem statement

As a neuropsychiatric disorder, Attention Deficit and Hyperactivity (ADHD) onsets on childhood inducing learning problems, low self-esteem, and substance use in adolescence.[1] Usually, ADHD diagnosis relies on information collected by family members, teachers, and clinical observation. However, the analysis depends on the specialist expertise, leading to high misdiagnosis rates.[2] To find a more specific and sensitive diagnosis, strategies integrating biomarkers and regular scientific assessments are required.[3]

In an attempt to describe reliable biomarkers for ADHD identification, current researches rely on the brain physiological response assessed through electroencephalography (EEG) to better understand neuropathologies.[4] For instance, the *Theta-Beta* power ratio (TBR) contrasts slow waves (4-7 Hz) and fast waves (13-30 Hz), under the assumption of more powerful slow waves in ADHD than in controls on resting state. Another proposed physiologically interpretable EEG feature corresponds to the P300 that expects a latency larger than the typical 300 ms after a stimulus is presented with lower amplitudes in ADHD subjects. Also, the error-related negativity (ERN) wave, evoked between 50-100 milliseconds at the frontal-central region when incorrect responses occur, has been considered to characterize ADHD children based on the impulsivity symptom.[5] Despite their physiological ground, recent works challenge the robustness and significance of TBR, P300, and ERN as ADHD biomarkers under the non-stationary and non-linear temporal dependencies of the EEG.[6–8]

In order to cope with temporal dependencies, stochastic tools are more appropriate. In particular, stochastic models unravel the temporal dynamics of time series by approximating the likelihood of assuming a parametric distribution over the observed signal.[9] A motivation of great theoretical and practical interest is to find a suitable similarity measure since these models can be a useful tool in various applications with multidimensional time series. One considered approach is the Kullback-Leibler (KL) divergence between probabilistic models to feed distance-based classifiers.[10] However, sometimes the KL divergence lacks

a closed-form solution, resulting in expensive Monte Carlo approximations.[11] Besides, the KL divergence does not satisfy the triangular inequality and symmetry properties required for the distance that parameterizes the classifier.[12] In contrast, the stationary cumulative distribution function fulfills the distance properties. Although the cumulative functions can be directly compared in the observation space,[13] it is designed for one-dimensional observations only .[14] Although stochastic models provide robust and efficient solutions, they have high computational costs so that computing similarity between two probabilistic models can be prohibitively intensive when applying these similarity measures to multichannel data.[15]

The success of stochastic models lies in the input information quality used in the model fitting. Thus, the use of relevant characteristics will allow better performance in the classification task. From the point of view of electrophysiological dependencies, it has been shown that the spectral features extracted from frequency bands such as *Delta*, *Theta*, *Alpha* and *Beta* a show a greater relevance in the discrimination of ADHD. For example, Ahmadlou et al.[4] combined the *Theta* band of the frontal and O2/P4 electrodes with the *Delta* band of the T5 electrode, getting a maximum accuracy of 95.6%. Sadatnezhad et al.[16] achieved a diagnostic accuracy of 86.4% from a combination of spectral power and fractal characteristics of EEG time series. Similarly, Abibullaev et al.[17] got relative *Theta* band measurements from recordings from nine frontal electrodes, achieving an accuracy of 97%. Therefore, diagnostic tools based on frequency dependencies of EEG recordings are promising, but such studies are still scarce, and the results do not offer a simple interpretation, so requires further replication and validation in this field of research.[18]

Despite the importance of a diagnostic support tool, interpretability is the key to effective decision support.[19] There is indeed great satisfaction when a considerable value in prediction accuracy is achieved, but it is a limitation when it is hardly interpretable. As an example, model complex has showed the ability to get very accurate diagnoses but in practice does not allow for an interpretation of the results.[20] Thus, despite their potential, there are still significant challenges to be addressed in clinical application.[21] In addition, in several cases the accuracy results are very fragile because they may solve the task inadequately, learning perturbations that can have a great impact on their outcome.[22] Thus, by explaining the reasoning behind the results, interpretable machine learning systems give users reasons to accept or reject the predictions.[23] Therefore, in diagnosis, the interpretability of a model is a quality that is imperative to achieve, as it ensures that the model can infer a patient's condition with complete justifications and the model can lead to a better understanding of the disorder.[24]

Thus, a strategy that models the temporal dependencies of EEG signals employing stochastic models, finding similarities in discriminative characteristics with a low computational cost and interpretable results at the clinical level for the diagnostic support of ADHD, can favor both an early screening and early interventions that reduce the degradation of the pathology in patients. Thus, the following research question is established: How to develop and implement a methodology to support ADHD diagnosis from the temporal

dynamics of multichannel EEG signals using a stochastic model that considers spectral information, reduced training cost, and interpretable results to facilitate decision making and reduce diagnostic errors?.

2.2 Justification

Attention deficit hyperactivity disorder (ADHD) can have a significant impact on patients' lives and is often associated with symptoms of inattention, hyperactivity, and impulsivity that vary in severity. Also, combined with the frequent presence of psychiatric comorbidities makes the assessment and diagnosis of ADHD challenging and can lead to treatment delays and wrong evaluations, resulting in difficulties in school,[25] work,[26,27] and interpersonal relationships.[28] ADHD symptoms can also increase risk behaviors, such as low self-esteem, suicidal gestures, psychoactive substance use, aggressiveness, and violence,[29] and therefore, have a high social impact because it is closely associated with impulsivity. Colombia has a high prevalence, particularly in Antioquia and the Eje Cafetero (15% to 17%), highlighting that of the three subtypes of ADHD (predominantly impulsive-hyperactive, inattentive, or combined), the combined and impulsive-hyperactive subtypes predominate in the Colombian population, unlike in industrialized countries where the impulsive-hyperactive type is the least frequent.[30,31]

At the economic level, Mental illness is increasingly seen as a global priority not only for health but also for global development.[32] One of the causes of the prioritization is their costs of morbidity, unemployment, and work absenteeism estimated at US$16 trillion over the next twenty years globally.[33] Colombia, for example, budgets to invest $\$1,120,850$ million pesos between 2020 and 2024 in improving the mental health of its citizens.[34]

At the administrative level, one of the main factors related to the deterioration of mental health in the Colombian population is the poor articulation between national policies and territorial entities for the effective promotion of mental health and prevention of mental disorders. However, strategies have been proposed such as the Plan Decenal de Salud Pública 2012-2021 proposes as a priority dimension, within the aim "la Cero tolerancia con la discapacidad evitable", mental health with a component of prevention and comprehensive care for mental problems and disorders. Within this component, the aim is to strengthen institutional and community management to ensure comprehensive care for mental problems and disorders and associated events. Given our social characteristics and prevalence of mental illnesses, it is essential to develop diagnostic support and improvement assessment tools for pathologies associated with poor impulse control, e.g., ADHD.[35] In this vein, understanding the neurophysiological mechanisms of impulsivity is of vital importance for a better understanding and management of ADHD.[36]

At the health care level, the aim is to compensate for the shortcomings in mental health issues through early identification and timely health care for people with mental health problems or disorders, as this is essential to reduce the progression of the disease and

reduce the socio-economic impact. Optimally, the health personnel who make the first contact with an individual entering the system are the most likely to refer a patient with mental health problems. However, the low availability of human talent for specialized care and the lack of tools for effective diagnosis weaken the promotion, prevention, diagnosis, treatment, and rehabilitation of the patient.

Taking into account the proposed diagnosis, the development of early detection tools based on neurophysiological biomarkers would allow more effective therapeutic action in the prevention of ADHD chronification, thus improving the quality of life of patients from the early stages of their life course.[37] In this regard, several attempts have been made to describe reliable biomarkers for the diagnosis of ADHD, which provide objective diagnostic elements and are inexpensive to extract. Among the experimental methods to try to answer this question are electrophysiological studies, mainly those based on EEG.[18] Therefore, the development of a tool to support ADHD diagnosis from temporal dynamic analysis of EEG electrophysiological changes using discriminative and interpretable brain electrical activity features using stochastic models is a promising approach to address this challenge.

Additionally, since this research is articulated with the project "Support tool for the diagnosis of ADHD in children from multiple characteristics of brain electrical activity of EEG recordings" presented by the research group of Automatics with registration number 63051 in MinCienciaas. It is expected to contribute to the improvement of mental health in patients with ADHD from the development of a novel diagnostic support tool that allows the specialist to have the interpretation of the results, and according to the approaches of social promotion marked in the project contribute to the promotion and prevention of ADHD reducing stigmatization and discrimination in patients.

2.3 State of the art

Clinical observation for the diagnosis of attention deficit hyperactivity disorder ADHD provides difficulties such as diagnostic subjectivity according to the cultural context, discrepancies in the information provided to specialists and overlap of symptoms with other disorders.[38] In addition, ADHD describes patterns of inattention, impulsivity and/or hyperactivity with diverse and context-dependent clinical presentations, which makes its identification, diagnosis and treatment difficult.[39,40] Therefore, several investigations have focused on promoting the search for diagnostic support tools that involve biomarkers to help the specialist reach a more specific and sensitive diagnosis of the pathology.[3]

Currently, among the proposed biomarkers there is great interest in assessing the morphological characteristics of the p300 wave through Cognitive Evoked Potential analysis. This wave is a marker of an attentional process through a positive deflection that appears approximately 300 milliseconds after a stimulus in stimulus discrimination tasks.[41] However, several clinical studies do not identify significant differences between ADHD patients and controls.[42] Another promising method, and the most published, has been the

Theta/Beta spectral ratio (TBN), approved by the US Food And Drug Administration (FDA) in June 2013, under the name Neuropsychiatric EEG-based Assessment Aid (NEBA).[43] This system continues to be debated by the international medical community because of the lack of experimental evidence on its robustness as a diagnostic tool. Among the limitations described are the lack of diagnostic sensitivity in the face of the great etiological, symptomatological and therapeutic response variability that is characteristic of patients with impulsivity disorders[44].[45] Also of interest are recent studies identifying the error-related negativity wave (or ERN/Ne) as a marker in patients with impulsivity disorders, which would be susceptible to changes according to emotional involvement in an inhibition task.[46]

In addition to the above two examples, a wide range of possible features extracted from EEG signals have been described in the literature with which it would be possible to discriminate between ADHD patients and control subjects: *Alpha*-independent diffuse slow activity with increased *Delta* and *Theta* (1-7 Hz), non-epileptiform focal slow activity, mixed fast and slow activity evidencing absence of *Alpha* and increased *Beta* frequency, excess of frontal dominant *Theta* activity or excess of *Alpha* frequency activity, frontal asymmetries, variable asymmetry, left greater than right or right greater than left, mainly in F3 F4 channels, increased *Alpha* activity in the temporal lobe, epileptiform activity, *Alpha* frequency greater than 12Hz in the posterior cortex, with normal to hyperbolic amplitudes, high *Beta* frequency with a morphological spindle, with anterior frequency, lack of appreciable *Alpha*-blocking with eye-opening, measures of brain connectivity, correlation dimension, Lyapunov exponents, approximation entropy, the fractal dimension, among others[40,45,47–49]

Of particular interest are recent studies identifying the error-related negativity waveform (or ERN/Ne) as an absent or low-amplitude marker in ADHD patients compared to control groups, and which would be susceptible to changes depending on emotional involvement in an inhibition task.[50] Studies discriminating between ADHD patients and control subjects from strategies based on some aforementioned EEG biomarkers and machine learning methods, such as support vector machines (SVM), deep learning or artificial neural networks,[51–53] have presented high performances in terms of classification sensitivities and specificities. For example, Ahmadlou and Adeli[4] present a novel multi-paradigm technique for EEG-based diagnosis of ADHD using an ingenious fusion of nonlinear science, wavelets, and neural networks. The signal is decomposed into its five sub-bands for each EEG channel and nonlinear features known as synchronization likelihood (SL) are extracted that quantifies the similarity of each electrode's signal with the others, finding that in ADHD subjects there are inefficient connections as it has lower SL. Likewise, Allahverdy et al.,[54] aiming to discriminate ADHD subjects, fed a Multilayer Perceptron (MLP) neural network with nonlinear EEG measures, such as Lyapunov exponent and Higuchi, Katz and Sevcik fractal dimensions. Achieving a classification accuracy for all electrodes of 68.6%, and accuracies in frontal, parietal, central, and occipital regions were 86%, 61%, 62%, and 55.6%, respectively. However, such strategies require further replication and validation. Also, many have problems of interpretability, which hinders their use as part of diagnostic

support tools.

Other studies have worked on the neural correlates of ADHD because they have the potential to improve diagnostic accuracy by indicating that altered neural connectivity is associated with the pathology in children. For example, Furlong et al.[55] used a sample of EEG signals in 52 unmedicated ADHD children and 77 typically developing controls where they measured connectivity as the synchronization of the time series of each pair of electrodes. The results showed that a hyperconnected neural network is associated with elevated ADHD symptom severity. On the other hand, Galindo et al.[56] obtained features from patterns of electrical activity sources from EEG recordings coming from a stop-signal paradigm, achieving an accuracy of 87% in the classification of subjects with the disorder. Despite these efforts, few studies consider the temporal dependencies of EEG in the diagnosis of ADHD. In applications to classify EEG, the most common algorithms are hidden Markov models (HMMs), as it can model the signal as a sequence of states.[57] Solhjoo et al.[58] published an investigation to classify mental tasks from modeling EEG recordings using HMM. The results reported accuracy of approximately 72% using multiple datasets. Moreover, Stam et al.[59] suggest that it associates brain activity states with chaotic dynamics nonlinear measures could quantify that such as entropy and Lyapunov exponents providing a robust model-independent, information-theoretic estimate of dynamical complexity.

2.4 Objectives

2.4.1 General objective

To develop a methodology to support the diagnosis of ADHD from multichannel EEG signals using stochastic models that taking into account signal dynamics, spectral information, computational cost reduction and inter- pretability of the results.

2.4.2 Specific objectives

- Develop a multichannel time series classification methodology taking into account signal dynamics by means of stochastic modeling.
- Develop a time series classification methodology that takes into account spectral information and reduces the computational cost of similarity.
- Develop a methodology for visualizing stochastic representations to facilitate the interpretability of inference machines.

Chapter 3

Develop a multichannel time series classification methodology taking into account signal dynamics

This chapter proposes a classification methodology that assesses the similarity between time-series using the probability product kernel (PPK) between their corresponding HMM. In this regard, PPK gathers the affinity of all jointly observable sequences through the Bhattacharya distance between two HMMs.[60] Since introduced PPK behaves as an inner product in the space of probability distributions, it feeds a support vector machine that performs the classification. We evaluate the proposed approach in the supported diagnosis of ADHD children, using EEG signals recorded while performing a cognitive paradigm magnifying the impulsivity symptom. The paradigm, termed Reward Stop-Signal Task (RSST), allows evaluating the classification results at three different rewards and two reward displaying orders. The comparison of the proposed PPK-based against the Log-likelihood-based features demonstrates that PPK improves the average classification accuracy while reducing the standard deviation.

3.1 Methods

3.1.1 Similarity between time series

Let a set of N time series $\mathcal{X}=\{\boldsymbol{X}_i \in \mathbb{R}^{D\times T_i}\}_{i=1}^{N}$ where each $\boldsymbol{X}_i=\{\boldsymbol{x}_{i0},\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT_i}\}$ corresponds to an ordered sequence of T_i vector-valued observations $\boldsymbol{x}_{it}\in\mathbb{R}^D$ belonging to a D-dimensional Euclidean space. A single Hidden Markov Model (HMM) describes the stochastic behavior of each $\boldsymbol{X}_i$ by finding the sequence of hidden states $\boldsymbol{q}_i=\{q_{i0},q_{i1},\ldots q_{iT_i}\}$ that maximizes the likelihood $p(\boldsymbol{X}_i|\boldsymbol{q}_i,\theta_i)$, with $q_{it}\in\{1,\ldots,M\}$ as one of M possible states. θ_i comprises the following parameters of the i-th HMM:

- The initial state probability distribution $\boldsymbol{\pi}=[\pi_m=p(q_0=m)]$, $\boldsymbol{\pi}\in[0,1]^M$ with constraint $\sum_{m=1}^{M}\pi_m=1$.
- The state transition probability distribution $\boldsymbol{A}=[a_{mn}=p(q_t=m|q_{t-1}=n)]$, $\boldsymbol{A}\in[0,1]^{M\times M}$, constrained by $\sum_{m=1}^{M}a_{mn}=1$.
- The emission probability density function that is considered as Gaussian $p(\boldsymbol{x}_t|q_t=m)=\mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m)$, for $m\in\{1,\ldots,M\}$.

To learn the parameter set θ_i, the Expectation-Maximization algorithm maximizes the likelihood in a two-step iterative way: The E-step estimates the most likely sequence of hidden states $\boldsymbol{q}_i$, and the M-step updates θ_i from the state posterior marginals.

Then, the similarity between a pair of time series $\boldsymbol{X}_i,\boldsymbol{X}_j$ can be assessed through the corresponding HMMs with parameters θ_i,θ_j using the probability product kernel between distributions (*PPK*):[60]

$$\kappa(\boldsymbol{X}_i,\boldsymbol{X}_j)\coloneqq\left\langle p(\boldsymbol{X}|\theta_i),p(\boldsymbol{X}|\theta_j)\right\rangle_{L_2}=\int p(\boldsymbol{X}|\theta_i)p(\boldsymbol{X}|\theta_j)d\boldsymbol{X} \tag{3.1}$$

where $\langle\cdot,\cdot\rangle_{L_2}$ stands for the inner product in the space of probability functions, and the integral accounts for all potential observable sequences.

Taking into account the factorization of the HMMs, Equation (3.1) can be approximated for a given sequence length T as in Algorithm 1, where $\psi_{mn}\in\mathbb{R}$ corresponds to the inner product between the observation pdfs from states m and n, which for two Gaussians $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m)$ and $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n)$ is computed in a closed form as:[61]

$$\begin{aligned}
\psi_{mn}&=\left\langle p(\boldsymbol{x}|q_m),p(\boldsymbol{x}|q_n)\right\rangle_{L_2}\\
\psi_{mn}&=\left\langle \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m),\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n)\right\rangle_{L_2}\\
\psi_{mn}&=(2\pi)^{(1-2\rho)D/2}\rho^{-D/2}\left|\boldsymbol{\Sigma}^{\dagger}\right|^{1/2}|\boldsymbol{\Sigma}_m|^{-\rho/2}|\boldsymbol{\Sigma}_n|^{-\rho/2}\\
&\qquad\times\exp\left(\frac{\rho}{2}\left(\boldsymbol{\mu}_m^{\top}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m+\boldsymbol{\mu}_n^{\top}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n-\boldsymbol{\mu}^{\dagger\top}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{\mu}^{\dagger}\right)\right)\\
\boldsymbol{\Sigma}^{\dagger}&=\left(\boldsymbol{\Sigma}_m^{-1}+\boldsymbol{\Sigma}_n^{-1}\right)^{-1}\\
\boldsymbol{\mu}^{\dagger}&=\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m+\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n
\end{aligned} \tag{3.2}$$

where $\boldsymbol{\Sigma}^{\dagger}$ stands for the pseudo-inverse of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ correspond the mean and covariance of the n-th Gaussian pdf, respectively.

Applying PPK at all HMMs pairs in $\mathcal{X}$ yields a symmetric definite positive kernel matrix $\boldsymbol{K}_{\mathcal{X}}(M,T)\in\mathbb{R}^{N\times N}$ with elements $k_{ij}(M,T)$. Note that the similarity measure between time series through their describing HMMs depends on two free parameters, namely, the number of discrete hidden states M considered to learn the HMM, and the sequence length T up to which the models are compared.

Algorithm 1 Probability Product Kernel for HMM with Gaussian emissions

Input: HMM parameters $\boldsymbol{\pi}^i, \boldsymbol{A}^i, \boldsymbol{\mu}^i_m, \boldsymbol{\Sigma}^i_m, \boldsymbol{\pi}^j, \boldsymbol{A}^j, \boldsymbol{\mu}^j_n, \boldsymbol{\Sigma}^j_n$ and PPK parameter T
Output: PPK approximation $k_{ij} \approx \kappa(\boldsymbol{X}_i, \boldsymbol{X}_j)$ between time series $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$

$$\varphi^0_{mn} = \left(\pi^i_m \pi^j_n\right)^{1/2} \quad \forall n,m \in \{1,\ldots,M\}$$

for $t = 1 \ldots T$ **do**

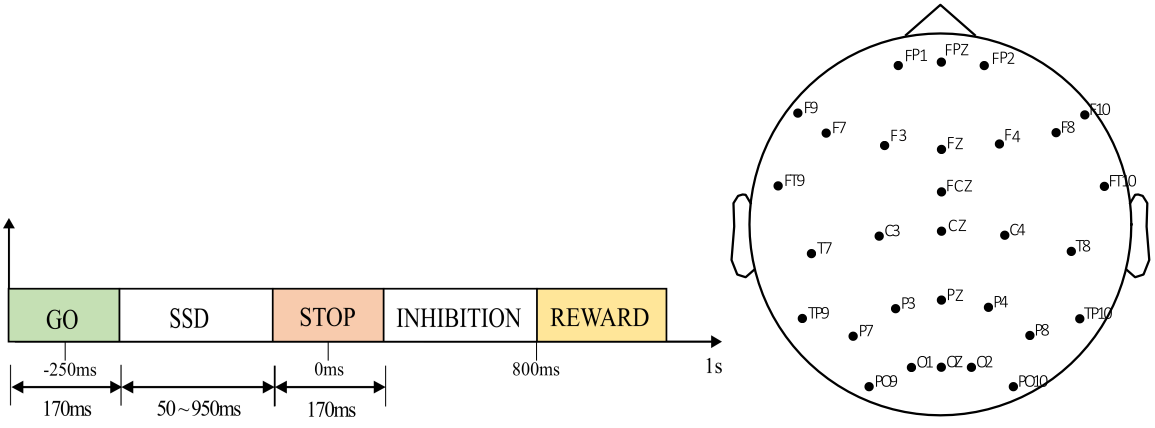
$$\varphi^t_{mn} = \sum_{m'=1}^{M} \sum_{n'=1}^{M} \psi_{m'n'} \left(a^i_{mm'} a^j_{nn'}\right)^{1/2} \varphi^{t-1}_{m'n'} \quad \forall n,m \in \{1,\ldots,M\}$$

$$k_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{M} \psi_{mn} \varphi^T_{mn}$$

3.2 Experimental Setup

3.2.1 EEG Dataset

The considered dataset holds EEG recordings from 63 children labeled as either healthy control (HC) or attention deficit hyperactivity disorder (ADHD) patient. The signal recording relies on a computerized executive inhibition paradigm, known as the Stop Signal Task (SST). Participants must press a key each time they face a frequent stimulus, called *Go*, except if a rare stimulus, called *Stop*, appears after the *Go*. Each time the *Go* appears on the screen defines a trial. If the participant inhibited the key pressing in a Go-Stop trail sequence, the paradigm labels the trial as successful inhibition; otherwise, as failed. The time interval between *Go* and *Stop* stimulus, known as the *Stop Signal Delay* (SSD), ranges between $250ms - 1000ms$ according to the participant performance, ensuring a rate of successful inhibitions around 50% (see Figure 3.1a).



(a) RSST trial. (b) Channel locations.

Figure 3.1. EEG signal acquisition paradigm. Bold labeled channels denote the channels of interest subset.

This work uses a new version of the SST paradigm, termed Reward Stop-Signal Task (RSST), exploring the reward magnitude modulation after a successful inhibition. The considered RSST holds four four-minutes blocks and two reward presentation orders, named conditions. The Increasing Condition (IC) rewards the participant with a *Smiley* sticker in the first block. The second and third blocks reward the participant with a *Low* amount of candies, and the last one with a *High* amount. On the contrary, the Decreasing Condition (DC) starts with a *Smiley* reward, followed by *High* in the second and third blocks, and finishes with a *Low* reward. Those two conditions allow evaluating not only the response to the reward magnitude but also abrupt reward changes.

Table 3.1 summarizes the number of subjects per class and condition, along with the average number of successful and failed inhibitions. Note that participants cannot perform the RSST with both conditions. Hence, they belong to either the IC or DC group. This work only studies failed inhibition trials, given the hypothesis that ADHD patients vary their Error Related Negativity features (ERN).[62–64]

Condition	**Diagnosis**	**N**	**Inhibition**	**Block of reward**			
				Block1	**Block2**	**Block3**	**Block4**
DC	**ADHD**	**15**	**Successful**	16.6±04	15.9±04	15.8±06	14.8±05
			Failed	12.9±04	14.0±04	12.9±06	14.0±06
	HC	**14**	**Successful**	15.3±03	17.7±04	17.3±03	18.0±04
			Failed	09.0±02	08.1±03	08.8±04	07.3±04
IC	**ADHD**	**17**	**Successful**	15.3±06	14.5±05	16.3±05	16.1±05
			Failed	14.6±06	14.8±05	13.3±04	13.3±05
	HC	**17**	**Successful**	10.2±04	11.3±04	11.2±05	11.8±05
			Failed	10.0±07	09.8±05	9.23±06	8.11±05

Table 3.1. Number of children per condition and average successful and failed inhibitions per block.

The brain electrical activity was recorded at a sampling frequency of 250Hz on 32 channels distributed over the scalp following the standard 10-20 montage, as shown in Figure 3.1b. Besides in the objective two, we carry out experiments with eight channels of interest (COI) registering the frontal-central EEG activity, that is consistently related to differential activity between ADHD and healthy control (HC) children.[65,66] Firstly, each trial is trimmed 200 ms before and 800 ms after the *Go* stimulus, producing sequences of T_i=250 time instants. Then, we centered and normalized each channel regarding its mean and standard deviation on the pre-*Go* period as in Equation (3.3), where $\boldsymbol{E}_{ref}\{\cdot\}$ is the averaging operator over the time before the *Go*.

$$\hat{\boldsymbol{x}}_{it} = \frac{\boldsymbol{x}_{it} - \boldsymbol{E}_{ref}\{\boldsymbol{x}_{it}\}}{\sqrt{\boldsymbol{E}_{ref}\left\{\left(\boldsymbol{x}_{it} - \boldsymbol{E}_{ref}\{\boldsymbol{x}_{it}\}\right)^2\right\}}} \tag{3.3}$$

3.2.2 HMM training

Since the RSST paradigm considers *Smiley*, *Low*, and *High* rewards, we train a single HMM for each of them, yielding three models per subject. In order to avoid overfitting, Expectation-Maximization (EM) algorithm learns HMM parameters using 70% of the failed inhibitions, and the remaining 30% stops the training. Figure 3.2, exemplifying the EM learning, evidences that the log-likelihood monotonically increases for the training EEG. On the contrary, the test data curve reaches a maximum at a few iterations, from which the model overfits the training data. Therefore, the hold-out appearance in HMM training not only guarantees the generalization of subject data but also reduces the number of EM iterations.

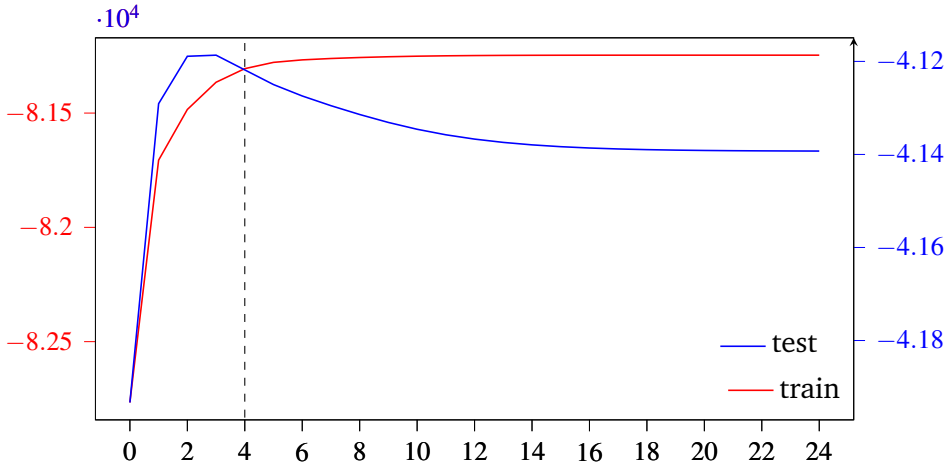


Figure 3.2. Log-likelihood along the EM iterations in the HMM training process for a control subject at *High* reward and M=3 states.

3.2.3 Parameter tuning and Classification

Given the set of trained HMMs, a support vector machine (SVM) fed by the probability product kernel carries out the classification of subjects into control or ADHD. Taking into account that the proposed methodology depends on the number of states, the sequence length, and the SVM box constraint, we tune such parameters using a 5-fold cross-validated grid search within $M \in \{3,\ldots,10\}$, $T \in \{2,\cdots,10\}$, and $C \in \{10^{-3},\cdots,10^{2}\}$.

3.3 Results

Figure 3.3 illustrates the attained classification accuracy averaged over the five test sets along the $M-T$ grid at the optimal box constraint for each reward and condition. For the

parameter tuning, the proposed approach achieves the best performance at M=6, T=2, and *High* with 90.0% accuracy, and at M=3, T=2, and *Low* with 71.0% for DC and IC children, respectively. Note that for both conditions, the larger the sequence length, the worse the accuracy because all the kernel values decrease, difficulting the classifier training. On the other hand, increasing the number of states reduces the accuracy at IC, while reaching a maximum performance at DC. Hence, gaining HMM complexity benefits the decreasing condition and hampers the increasing one.

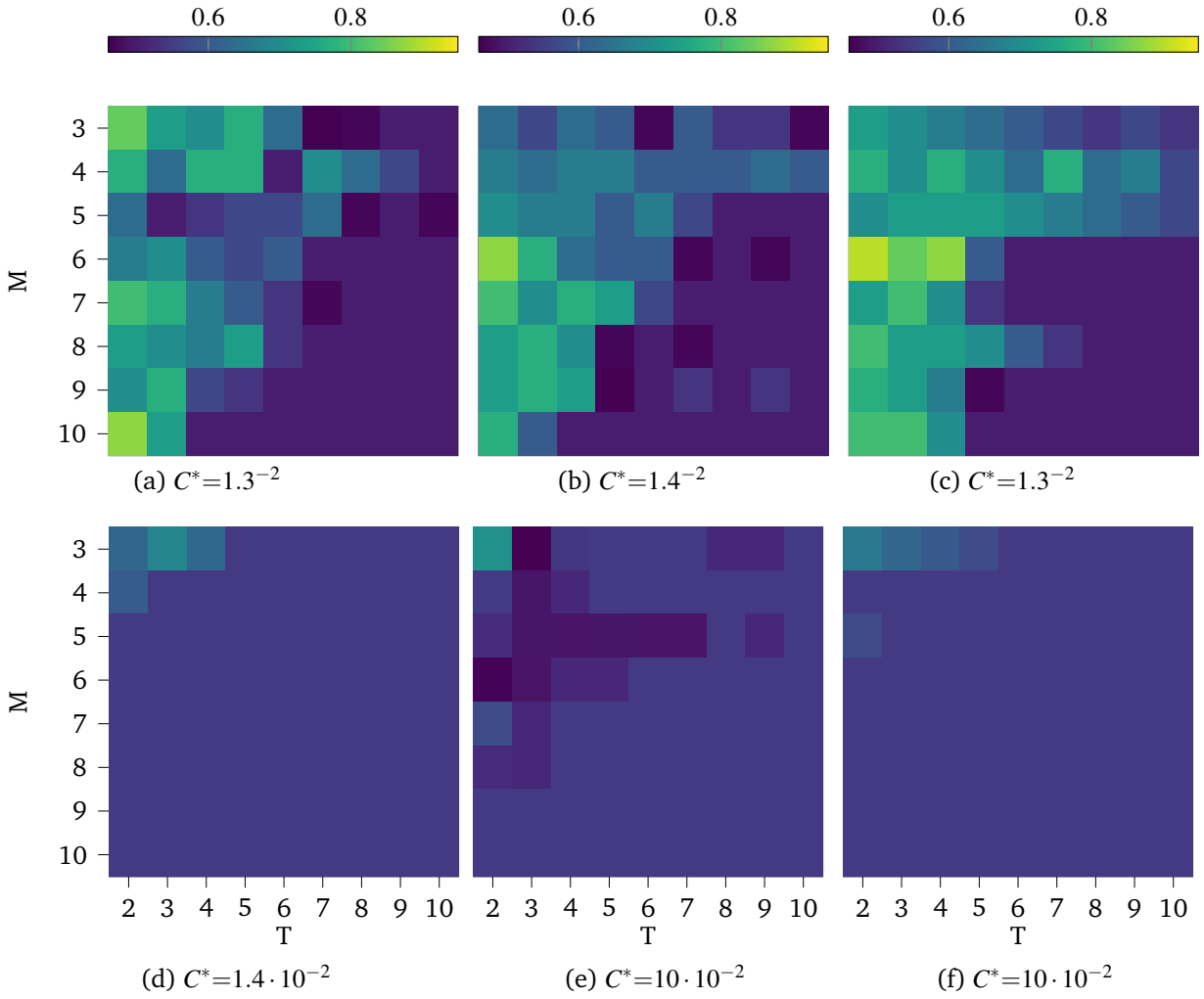


Figure 3.3. PPK classification accuracy for each evaluated parameter set at decreasing (top) and increasing (bottom) conditions. Accuracy is averaged over five test folds. Left to right: *Smiley*, *Low*, and *High* rewards.

Since the proposed methodology matches the time series in terms of their HMM, we compare against the classical HMM-based classification scenario, where a single HMM models each class and reward.[67] Then, an SVM classifies a target subject using as features

the log-likelihoods (LL) of the subject trials given the HMM of their corresponding reward. We note that SVM parameters for LL were trained in the same PPK training scheme, using a Gaussian kernel.

Table 3.2 summarizes the average accuracy with standard deviation attained by PPK and LL for considered conditions and rewards. Results evidence that the PPK approach outperforms LL at all rewards and conditions. Further, the more significant accuracies on DC than on IC prove that such a condition is more adequated for supported diagnosis purposes. Regarding the reward-condition pair, *High* at DC attains the best performance with a suitable standard deviation, which agrees with the RSST paradigm designed to highlight differences between ADHD and controls. As a result, the introduced PPK enhances the supported diagnosis of ADHD by assessing the similarity between the electrical activity of subjects.

		Reward		
Condition	*Approach*	*Smiley*	*Low*	*High*
DC	*LL*	73.3 ± 9	68.0 ± 7	64.4 ± 11
	PPK	$\mathbf{87.0 \pm 13}$	$\mathbf{87.0 \pm 7}$	$\mathbf{90.0 \pm 8}$
IC	*LL*	68.1 ± 17	62.8 ± 21	67.8 ± 20
	PPK	68.0 ± 12	71.0 ± 8	65.0 ± 8

Table 3.2. Classification results for PPK and LL per conditions and rewards.

3.4 Discussion

The present work aims develop a multichannel time series classification methodology taking into account EEG signal dynamics to ADHD diagnosis. The proposed methodology assesses the similarity between EEG signals using the PPK between their corresponding HMMs. Thanks to the HMM factorization, PPK can be approximated in a closed form for a given sequence length and number of hidden states. Therefore, the resulting kernel can feed classifiers based on inner-products, such as a support vector machine.

To test the methodology, we consider EEG signals recorded under the RSST paradigm at three reward levels and two ways to present rewards, termed conditions, yielding six EEG subsets. Since PPK depends on the number of states and sequence length, a cross-validation scheme searches for the highest accuracy in a parameter grid. Tuning results in Figure 3.3 shows that performance decreases as the sequences become longer. Further, either a few or a large number of states hinders the differences among subjects, so that there exist an intermediate M unraveling class differences.

Comparison results in Table 3.2 show that PPK outperforms LL at the decreasing condition while behaving similarly at the increasing one. Besides, since PPK deviations are shorter than LL, the proposed approach yields a more confident classification performance.

Regarding the RSST paradigm, we only considered failed inhibitions trials relying on the executive dysfunction and the deregulation of the inhibitory functions due to ADHD. From attained accuracies, we identified *High* as the best reward for DC condition, agreeing with the motivational changes associated with the attention deficit.[68]

Chapter 4

Develop a time series classification methodology that takes into account spectral information and reduces the computational cost of training.

In this chapter, we propose a similar kernel-based methodology with reduced computational cost to facilitate the interpretation and influence of M, T, and *rho* parameters in the classification task of HMMs trained with EEG recordings. The proposed method streamlines the calculation of PPK for HMMs by allowing more flexibility in parameter tuning. This chapter, not only analyzed the effect of sequence length T and the number of states M but also the power ρ. To carry out the above, we created two synthetic databases, a bi-class of sequences generated by two HMMs with different transitions to identify the influence of the parameters on the similarity kernel and a multiclass with sequences generated by three types of attractors to identify the effect of the parameters on the classification stage. To answer our general aim, we used the RSST database being stricter in the paradigm's execution, so we used the EEG records of each block associated with a reward level. The results are presented in two-channel configurations comparing raw and filtered EEG recordings in *Theta*, *Alpha*, and *Beta* bands. As a result, we identified that tuning parameters together with channel selection in a region of interest and separation of EEG into rhythms allow for better ADHD classification performance.

4.1 Methods

4.1.1 Fast computation of PPK for HMM

Despite integrating over all observable sequence is untractable, the HMM factorization property allows approximating Equation (3.1) for a given number T of transitions between states of θ_i and θ_j as:

$$\kappa\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) \approx k\left(\theta_i, \theta_j\right) = \sum_{m'=1}^{M} \sum_{n'=1}^{M} \left(\pi_m^i \pi_n^i\right)^{\rho} \psi_{mn} \prod_{t=1}^{T} \left(a_{mm'}^i a_{nn'}^j\right)^{\rho} \psi_{m'n'} \tag{4.1}$$

Since the number of operations to compute PPK according to Equation (4.1) linearly grows with T, a practical implementation becomes computationally intensive for large time series datasets.[69] This work rewrites the PPK estimation in a matrix closed form reducing the computation time. To this end, let the similarity matrix between all Gaussian observation pairs from θ_i and θ_j as:

$$\boldsymbol{\Psi} = \{\psi_{mn} : \forall m, n\} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{m1} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{1n} & \psi_{2n} & \cdots & \psi_{MN} \end{bmatrix} \tag{4.2}$$

Then, the vector $\boldsymbol{\alpha}_0 \in [0,1]^{1 \times MN}$ in Equation (4.3) holds all the products between the initial state and observation probabilities:

$$\boldsymbol{\alpha}_0 = \underbrace{\left(\boldsymbol{\pi}^i \otimes \boldsymbol{\pi}^j\right)^{\rho} \odot \operatorname{vec}\left(\boldsymbol{\Psi}\right)}_{1 \times MN} \tag{4.3}$$

Where $\otimes$ and $\odot$ denote the Kronecker and Hadamard products, respectively; and $\operatorname{vec}(\cdot)$ stands for the operator vector. Similarly, Equation (4.4) calculates the product between transition and observations probabilities:

$$\boldsymbol{\alpha} = \underbrace{\left(\boldsymbol{A}^i \otimes \boldsymbol{A}^j\right)^{\rho} \odot \left[\operatorname{vec}\left(\boldsymbol{\Psi}\right), \operatorname{vec}\left(\boldsymbol{\Psi}\right), \ldots, \operatorname{vec}\left(\boldsymbol{\Psi}\right)\right]}_{MN \times MN} \tag{4.4}$$

Since the matrix $\boldsymbol{\alpha}$ comprises all the possible combinations between the Gaussian emissions of two HMMs through one-step transitions, the calculation for T-steps is achieved as $\boldsymbol{\alpha}^T$. Lastly, the initial similarity vector and the transition similarity matrix are joined to assess the PPK similarity as:

$$\kappa\left(\theta_i, \theta_j\right) = \operatorname{sum}\left(\boldsymbol{\alpha}_0 \boldsymbol{\alpha}^T\right) \tag{4.5}$$

The application of Equation (4.5) to $\mathcal{X}$ results in a kernel matrix $\boldsymbol{K}_\theta \in \mathbb{R}^{N\times N}$ holding all pair-wise time-series similarities which can be fed into any inner product-based pattern recognition machine such as a support vector machine for classification.

4.2 Experimental Setup

4.2.1 Synthetic Database

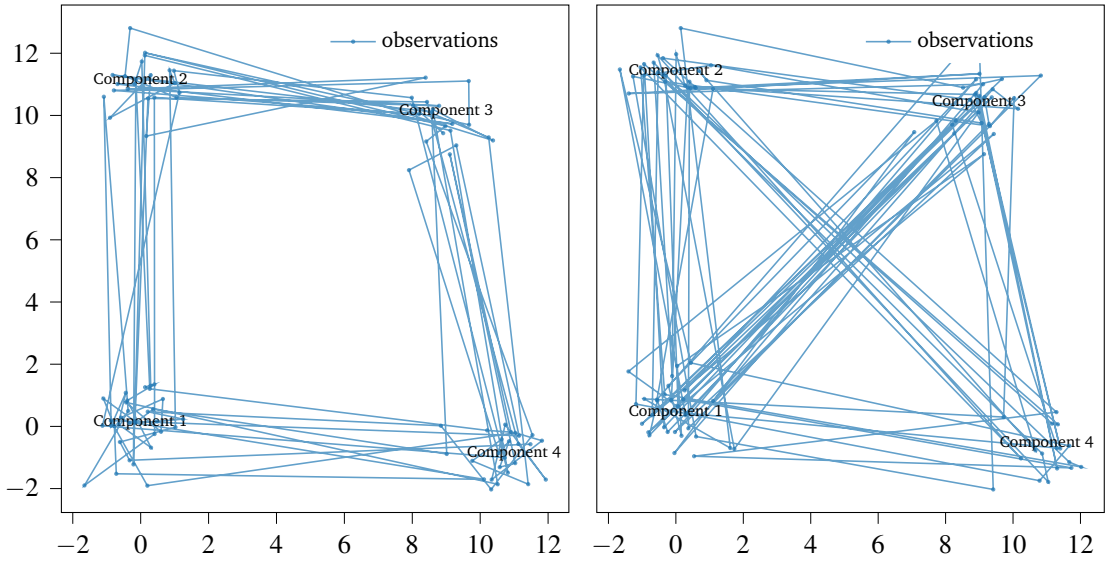
To identify the role of the parameters M, T and ρ in the similarity kernel, we created a synthetic biclass and a multiclass database. In the first one we generated fifty sequences of $\boldsymbol{X}$=$\{x_0, x_1, \cdots, x_{250}\}$ samples from two four-state HMMs, one with crossed transitions and the other with circular transitions but with equal emission probability density and priors as shown Equation (4.6).

$$\boldsymbol{A_i} = \begin{bmatrix} 0.2 & 0.3 & 0.3 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.3 \\ 0.4 & 0.3 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.3 \end{bmatrix} \qquad \boldsymbol{A_j} = \begin{bmatrix} 0.7 & 0.2 & 0.0 & 0.1 \\ 0.3 & 0.5 & 0.2 & 0.0 \\ 0.0 & 0.3 & 0.5 & 0.2 \\ 0.2 & 0.0 & 0.2 & 0.6 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 11.0 \\ 9.0 & 10.0 \\ 11.0 & -1.0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}^4 \tag{4.6}$$

$$\boldsymbol{\pi} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

The Figure 4.1 evidences the sequence of observations generated with the transitions between them. We can see that, in Figure 4.1a as specified in our transition matrix, there is no transition between states one and three, but in the second model there are transitions between all states as see Figure 4.1b.



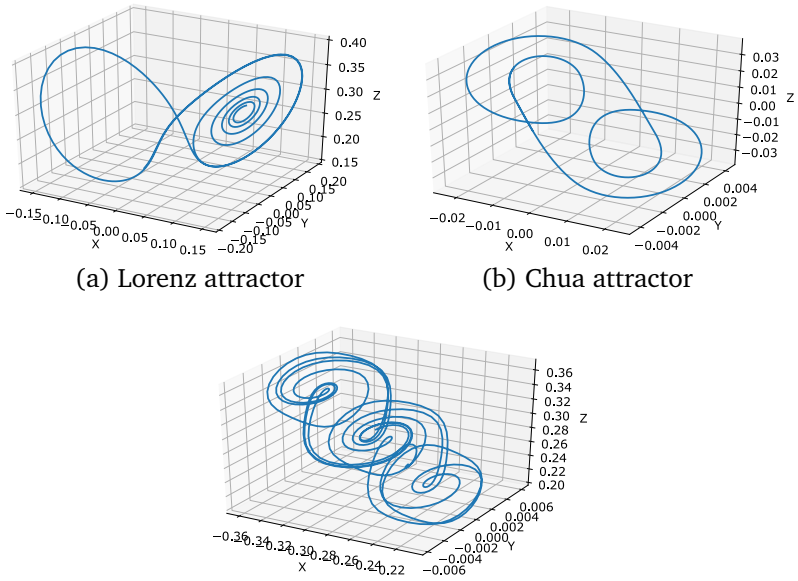
(a) circular transitions (b) cross transitions

Figure 4.1. Sampling from two HMM with circular and crossed type transitions.

Taking into account the influence of parameter tuning on the classification task. We generate a multi-class database with fifty sequences of $\boldsymbol{X}_i \in \mathbb{R}^{D \times T_i}$ where each $\boldsymbol{X}_i = \{\boldsymbol{x}_{i0}, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i500}\}$ obtained by three types of attractors as visualized in the Figure 4.2. This database is considered because the attractors have a 3D representation, so we expect that the methodology will find the best set of parameters that favor the separability of the classes through the dynamic evolution of the data and not on static component methodologies.

4.2.2 Training and Parameter tuning and classification

The proposed methodology comprises three stages: The first one trains a subject-wise HMM using the EEG trials from failed inhibitions. The second stage computes the PPK among HMMs. The last one learns a support vector machine (SVM) classifier from the precomputed kernel. Note that the full proposal depends on the number of states, the number of transitions, the PPK power, and the box constraint. Therefore, we tune those four hyperparameters using a five-fold cross-validated grid search for the highest classification accuracy within $M \in [3, 10]$, $T \in [1, 10]$, $\rho \in [0.1, 1.0]$, and $C \in [10^{-3}, 10^{2}]$.To discriminate ADHD and HC children, we evaluate the proposed methodology for the two conditions, four blocks, and two EEG channel setups. For the second and third stages, the methodology computes the pair-wise subject kernel and trains an SVM using HMMs from the same condition, block, and channel setup. Additionally, we carry out a similar classification with a radial basis function (RBF) kernel for feed an SVM using the behavior data (BD) of Table 3.1, where



(a) Lorenz attractor

(b) Chua attractor

(c) Multi-scroll attractor

Figure 4.2. Multi-class database from signals acquired from three different attractors.

each subject has eight characteristics; the number of successful and unsuccessful inhibitions in the four blocks. Also, due to the decomposition of brain electrical activity into frequency rhythms with a physiological interpretation, we carry out a spectral analysis that aims to provide insights into the underlying ADHD mechanisms. To this end, we apply the supported diagnosis methodology from Section 4.2 to the EEG dataset band-pass filtered within the following brain rhythms: Theta [4,8]Hz, Alpha [8,12]Hz, and Beta [13,30]Hz.

4.3 Results

Figure 4.3 shows the effect of the parameters $M-T-\rho$ on the average similarity calculation of the first, second, and fourth quadrant of the kernel. For this purpose, we used two parameter settings $M=3, T=1, \rho=0.1$ and $M=10, T=10, \rho=1.0$, which allowed us to identify that as the number of states increases there are no significant changes in the similarity value for the three quadrants of the kernel. However, increasing the sequence length and ρ power along the number of states evidences a less average similarity in quadrant two, which represents a greater separability between the two classes. Conversely, increasing sequence length or ρ power shows changes in similarity in all three quadrants. Note that similarity in the second quadrant decreases rapidly as the other parameters increase, while similarity in the first and fourth quadrants decreases more slowly in comparison when we

have small values.

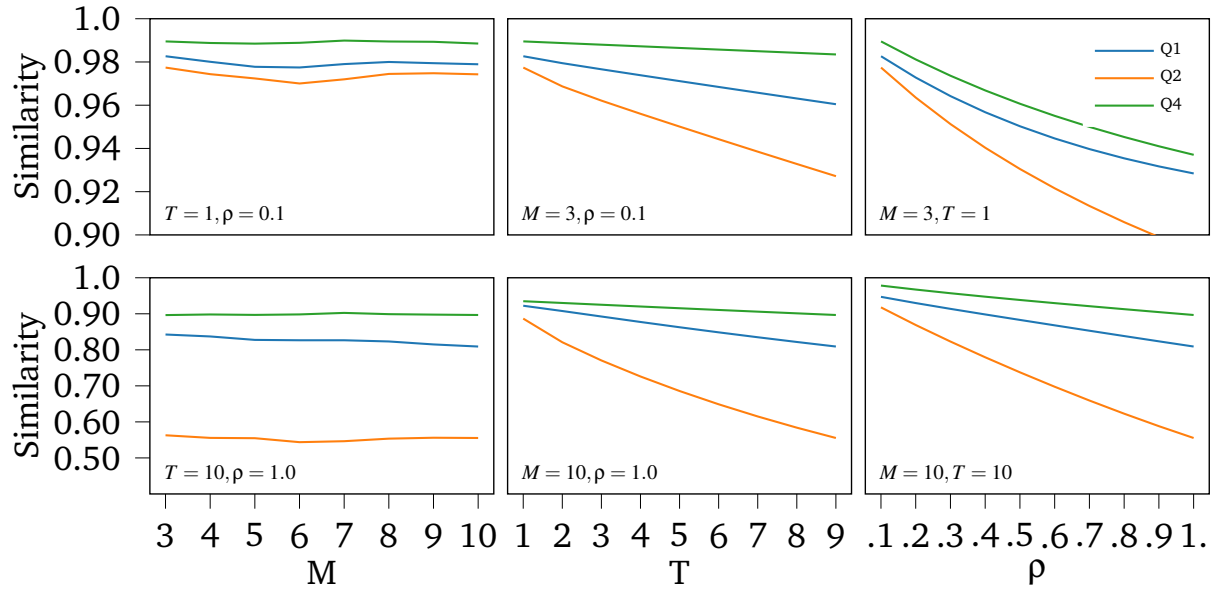


Figure 4.3. Effect of M, T, and ρ parameters on the average of each quadrant of the similarity kernel.

Figure 4.4 shows the parameter tuning in the similarity kernel under a main parameter configuration $M = 6$ - $T = 16$ - $\rho = 0.3$. The Figure 4.4c evidences the changes in average precision along with T and ρ, noting that when both parameters are increased, the precision decreases. Similary, the Figure 4.4a shows that increasing the model complexity and using small ρ increases the accuracy. The opposite occurs in Figure 4.4b, where increasing the number of states and the length of the long sequence improves the classification performance.

Figure 4.5 visualizes the target kernel, the Bhattacharya kernel, and the similarity kernel with parameter tuning of the multiclass classification task. Figures 4.5b and 4.5c presents a chess effect in the second and third class similarity that can be explained from the modes of the signal distribution although it does not present any impediment to reach a high accuracy value. Also, given the characteristics of the last quadrants of both Kernel, we could infer it is difficult to discriminate the sequences of the second and third class. Despite this, the average accuracy was 99.3% for the Kernel with $\rho = 0.3$ and the Bhattacharya kernel 98%.

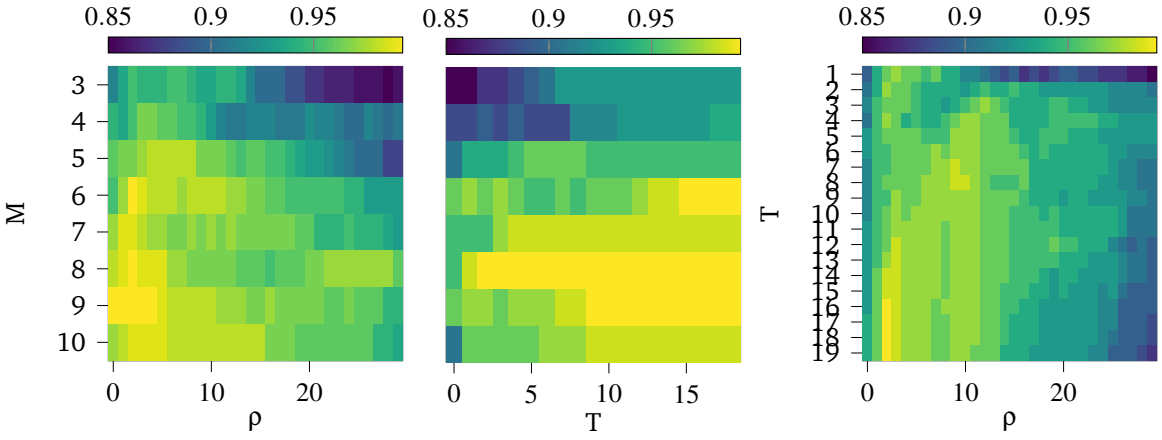


Figure 4.4. Effect of M, T, and ρ parameters on the classification task of the multi-class database.

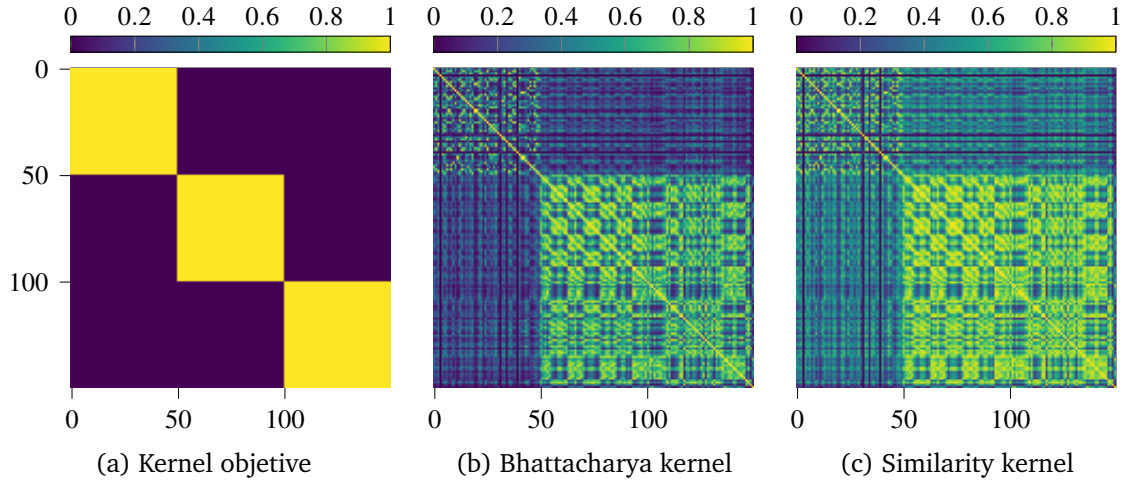
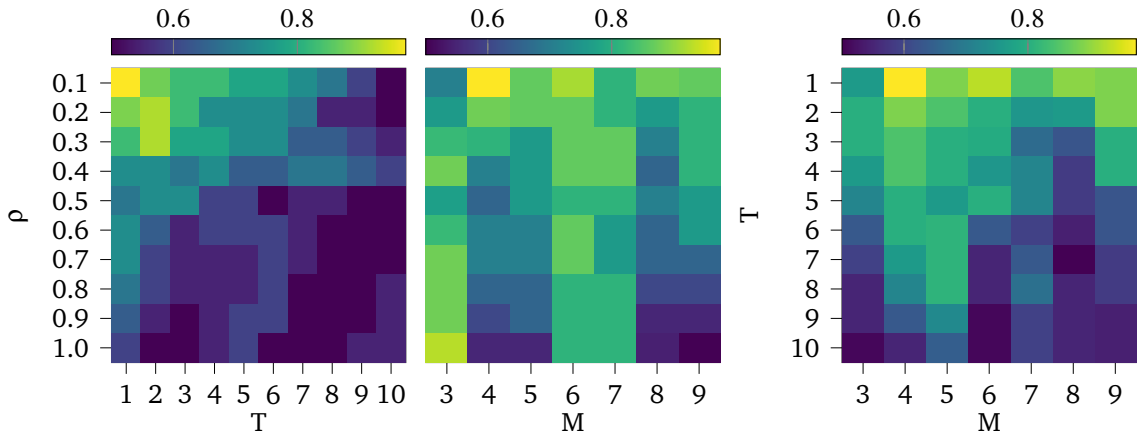


Figure 4.5. Effect of parameters M, T and *rho* on the similarity kernel.

Figure 4.6 presents the average validation accuracy along the number of states, the number of transitions, and power values. It is worth noting that an excessive increment on the PPK parameters dramatically decreases the classification accuracy. On the contrary, the number of HMM states lacks a monotonical influence. Therefore, the span of M to find the relevant performances becomes larger.



(a) $M = 4$ (b) $T = 1$ (c) $\rho = 0.1$

Figure 4.6. Validation accuracy averaged over five test folds, for each evaluated parameter set at DC, Block 1, and COI.

Table 4.1 presents the optimal hyperparameters for each condition, block, and channel setup, along with the attained accuracy and F1-score. Regarding the channel setup, HMMs trained with COI outperform the 32 channels montage. From the point of view of behavior data, the results of classification suggests that under the influence of RSST paradigm, ADHD and control subjects respond in different manners when confronted to an inhibitory task.Consequently, the frontal-central area highlights electrical activity related to different impulse control modulations in the presence of rewards, as a key element for ADHD diagnosis. With respect to the RSST paradigm, the decreasing condition obtains higher classification scores than increasing, with the best results at the first two blocks. Further, the reached score of 93% implies that ADHD and HC subjects differently respond when confronted with an inhibitory task, so that the stochastic dynamic of EEG becomes significantly discriminative.

Ch	*C*	*B*	*M*	*T*	ρ	*Acc*	*F1*
BD	DC					86±13	87±13
	IC					88±15	85±02
All	DC	1	6	1	0.5	80±12	84±09
		2	**6**	**1**	**0.1**	**83±10**	**86±08**
		3	10	1	0.6	73±13	80±07
		4	6	1	0.1	77±13	82±10
	IC	1	**6**	**1**	**0.1**	**80±06**	**84±08**
		2	3	1	0.1	65±14	71±08
		3	6	1	0.1	60±16	52±30
		4	5	1	0.1	77±14	82±11
COI	DC	1	**4**	**1**	**0.1**	**93±08**	**93±13**
		2	7	1	0.7	76±13	80±11
		3	3	1	0.3	90±13	91±13
		4	4	1	0.5	90±13	91±13
	IC	1	4	5	0.1	91±11	91±13
		2	3	2	0.2	86±09	85±11
		3	4	3	0.3	83±10	83±12
		4	**5**	**1**	**0.2**	**91±11**	**92±10**

Table 4.1. Classification results per condition (*C*), block (*B*), and channel setup (*Ch*). Average and standard deviation of five folds is presented for the accuracy and F1 score.

Table 4.2 displays the supported diagnosis performance for the considered frequency bands at each block and condition with the optimal hyperparameters. Results evidence different block performances within reward order conditions. For DC, the first block consistently reaches the highest F1 score, followed by the third one. Conversely, the most discriminative block varies from rhythm to rhythm for IC. However, the Beta rhythm always exhibits the highest performance, reaching 97% accuracy for IC that improves the results without the spectral decomposition in Table 4.1.

C	*R*	*B*	*M*	*T*	ρ	*Acc*	*F1*
DC	Theta	1	**8**	**1**	**0.1**	**86±12**	**90±08**
		2	5	1	0.1	79±07	69±25
		3	5	1	0.4	86±13	79±13
		4	4	1	0.3	79±14	66±19
	Alpha	1	**3**	**2**	**0.1**	**90±08**	**84±09**
		2	3	1	0.5	82±16	77±21
		3	3	2	0.1	83±10	75±08
		4	7	3	0.1	83±16	72±09
	Beta	1	**4**	**1**	**0.7**	**90±08**	**93±09**
		2	5	5	0.2	79±16	79±12
		3	5	1	0.1	86±12	87±07
		4	3	1	0.3	79±20	61±37
IC	Theta	1	**4**	**4**	**0.1**	**94±07**	**84±23**
		2	5	1	0.6	97±06	80±16
		3	3	4	0.1	85±14	81±17
		4	5	1	0.5	94±10	83±11
	Alpha	1	7	1	0.1	88±15	69±38
		2	**3**	**5**	**0.3**	**88±10**	**80±14**
		3	5	1	0.1	85±08	67±16
		4	6	2	0.1	85±13	71±22
	Beta	1	**3**	**1**	**0.6**	**97±05**	**92±08**
		2	4	1	0.1	91±06	80±13
		3	**6**	**1**	**0.1**	**97±05**	**92±08**
		4	4	1	0.1	94±06	84±23

Table 4.2. Classification results per condition (*C*), block (*B*), and rhythm (*R*). Average and standard deviation of five folds is presented for the accuracy and F1 score.

4.4 Discussion

This chapter proposes a supported diagnostic methodology for ADHD children, based on kernel-similarity with reduced computational cost to facilitate the interpretation and influence of M, T, and ρ parameters in the classification task of HMMs trained with EEG recordings using two configurations of channels and spectral analysis.

Through the calculation of the PPK fast, we checked the influence of the sequence length, the ρ power, and the number of hidden states on the classification, which allowed us to pinpoint the need to tune these parameters to achieve maximum success. Also, we

showed that the effect of the parameters varies according to the characteristics of each database. For example, in the bi-class database increasing the number of states is not significant but increasing the sequence length and the ρ power increases the separability between classes as shown Figure 4.3. Contrary to the multi-class database, where increasing the number of states and the sequence length increased the average accuracy for small values of ρ (see Figure 4.4). In the RSST database, increasing the number of states generates a greater complexity in the calculation of the similarity kernel, although having very small values of T and ρ, this difficulty is overcome. Likewise, when T and ρ increase, the accuracy decreases due to the effect of overfitting the similarity kernel learning the particularities of each subject and not those of the class as shown Figure 4.6.

To test the methodology, we considered EEG signals recorded under the RSST paradigm regarding of two conditions, three reward levels, and four blocks where each subject has four EEG recordings. With the aim of highlight electrophysiological changes according to the motivational effect in the classified ADHD and control subjects, Table 4.1 evidence that behavioral data are not sufficient as a diagnostic support tool evidence that our methodology exceeds the results with behavioral data, which agrees that neuropsychological tests do not allow correct discrimination of ADHD [70].[71]

Regarding the conditions, in IC we see a difference between controls and ADHD in both the first and last blocks. This may point at a distinct motivation allocation in both groups. The first block marks the fact that children engage in the task with differences that could rely on the effect of novelty, whereas different results in the last block, may reflect a distinct sensibility to the magnitude of the reward, given by the exposition to the higher reward at the end of the task.Likewise, in DC we also see the effect of novelty in the first block, similar to that observed in IC, but then we see that the accuracies fall in *block2* and *block3*. For the DC group, in these two blocks, the children are exposed to the highest reward. This result suggests that exposure to high reward induces a strong change in the brain electrical activity, to the point that it becomes difficult to distinguish such electrical activity between control children and ADHD [7273].[74]

Considering a frequency bands analysis, we perform filtering in *Theta*, *Beta*, and *Alpha* rhythms of the EEG signals in select channels for each block and condition, taking to account the RSST paradigm. Table 4.2 shows that in DC group, the first block exhibited the best performance over three bands, which is consistent with the results shown in Table 4.1. Although in IC this phenomenon does not occur, the results are superior to those obtained in the previous approaches.

Chapter 5

Develop a methodology for visualizing stochastic representations to facilitate the interpretability of inference machines

This chapter presents a methodology that allows spatially interpretation of HMMs through dimension reduction. With this aim, we achieve visualization of HMMs as a function of state models, where an inner product between pairs of probability distributions represents the kernel $\boldsymbol{K}_\theta$ projected in 2D space. On the other hand, we propose a visual representation of the time-series dynamics through the inner product between pairs of probability densities of each state $\boldsymbol{K}_\psi$, identifying each (x,y) coordinate in the representation space as a state of each HMM which represents a topoplot describing the mean vector of each state μ_{im}. Also, to identify dynamical changes, we use the transition probabilities a^i_{mn} to relate the brain activity changes of the topo-plots, so the more likely the transition, the more prominent the connection. As a result, spatial channel selection and spectral filtering change the distribution of EEG data and highlight their dynamics to improve class separability.

5.1 Methods

5.1.1 Model interpretability

In order to integrate the kernel operator functions into a Principal Component Analysis representation, consider a nonlinear transformation $\phi(x)$ into a D-dimensional feature space where de data x_i is projected in a point space $\phi(x_i)$ and that the projected data set has zero

mean. The covariance matrix in feature space is given by

$$\mathbf{C} = \frac{1}{l}\sum_{i=1}^{l} \phi(\mathbf{x}_i)\,\phi(\mathbf{x}_i)^{\mathrm{T}} \tag{5.1}$$

Taking into account the eigenvector expansion $\mathbf{CV} = \lambda\mathbf{V}$ and that the vector V is given by a linear combination of $\phi(x_n)$ as $\mathbf{V} = \sum_{i=1}^{l} a_i\phi(\mathbf{x}_i)$, we can substituting the ecuation Equation (5.1) as:

$$\frac{1}{l}\sum_{i=1}^{l} \phi(\mathbf{x}_i)\,\phi(\mathbf{x}_i)^{\mathrm{T}} \sum_{j=1}^{l} a_j\phi\left(\mathbf{x}_j\right) = \lambda\sum_{i=1}^{l} a_i\phi(\mathbf{x}_i) \tag{5.2}$$

Now expressing in terms of the kernel function $\kappa(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$, we arrive at $\ell\boldsymbol{\lambda}\mathbf{K}\mathbf{a} = \mathbf{K}^2\mathbf{a}$. Where $\mathbf{a}$ is a column vector with entries $a_1,\ldots,a_l$ and whose solution is $\ell\boldsymbol{\lambda}\mathbf{a} = \mathbf{K}\mathbf{a}$ allow that the eigenvectors in feature space be normalized is given by:

$$1 = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \sum_{i,j}^{\ell} a_i a_j\phi(\mathbf{x}_i)^{\mathrm{T}}\phi\left(\mathbf{x}_j\right) = \mathbf{a}^{\mathrm{T}}\mathbf{K}\mathbf{a} = \lambda\ell\mathbf{a}^{\mathrm{T}}\mathbf{a} \tag{5.3}$$

So far it has been assumed that the projected data has a zero mean, which is not guaranteed in the characteristics space. So the projected data after centralizing is given by $\tilde{\phi}(\mathbf{x}_n) = \phi(\mathbf{x}_n) - \frac{1}{N}\sum_{l=1}^{N}\phi(\mathbf{x}_l)$, which will allow to calculate $\widetilde{K}_{nm}$ as showing the Equation (5.4), where $\mathbf{1}_N$ denotes the $N\times N$ matrix in which every element takes the value $\frac{1}{N}$.

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N \tag{5.4}$$

5.1.2 Low-dimensional HMM visualization

In an attempt to visualize time series dynamics using similarity kernel, we introduce the Low-dimensional HMM visualization from Kernel Principal Component Analysis (KPCA) representation. Consider the matrix $\boldsymbol{K}_\theta \in \mathbb{R}^{N\times N}$, holding all pairwise between HMM θ_i of observations in $\mathcal{X}$, where each kernel element represents inner products between space of probability, as Equation (3.1). $\boldsymbol{K}_\theta$ corresponds to a Gram matrix with eigenvectors $\boldsymbol{\nu}_l \in \mathbb{R}^N$ spanning the space of observation distributions, that is:[?]

$$N\boldsymbol{K}_\theta\boldsymbol{\nu}_l = \lambda_l\boldsymbol{\nu}_l \tag{5.5}$$

With $\lambda_l \in \mathbb{R}^+$ as an eigenvalue of $\boldsymbol{K}_\theta$. Selecting the L largest eigenvalues yields the truncated matrix of eigenvectors $\boldsymbol{V} \in \mathbb{R}^{N\times L}$, where each row represents a state model in terms of a finite L-dimensional vector $\boldsymbol{v}_i \in \mathbb{R}^L$.

5.1.3 Low-dimensional state visualization

In order to endow the HMM with interpretability, we visualize the dynamics of the time series through the distribution probability densities of each state from Kernel Principal Component Analysis (KPCA). Consider the set of Gaussian distributions $\mathcal{G} = \{p_i(\boldsymbol{x}|q_m) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im}) : \forall i \in [1,N], m \in [1,M]\}$ modeling the observations in $\mathcal{X}$, where $p_i(\boldsymbol{x}|q_m)$ denotes the observation distribution for the i-th time series at state m. The matrix $\boldsymbol{K}_\Psi \in \mathbb{R}^{NM \times NM}$, holding all pairwise PPK between Gaussians as in Equation (4.2), where each kernel element represents inner products between Gaussian distributions, achieving a visualization of the model states. $\boldsymbol{K}_\Psi$ corresponds to a Gram matrix with eigenvectors $\mathbf{v}_l \in \mathbb{R}^{NM}$ spanning the space of the observation distributions:

$$NM\boldsymbol{K}_\Psi \mathbf{v}_l = \lambda_l \mathbf{v}_l \tag{5.6}$$

With $\lambda_l \in \mathbb{R}^+$ as an eigenvalue of $\boldsymbol{K}_\Psi$. Selecting the L largest eigenvalues yields the truncated matrix of eigenvectors $\boldsymbol{V} \in \mathbb{R}^{NM \times L}$, where each row represents a Gaussian distribution in $\mathcal{G}$ in terms of a finite L-dimensional vector $\boldsymbol{v}_{im} \in \mathbb{R}^L$.

For the purpose of visualization, $\{\boldsymbol{v}_{im} : \forall m\}$ corresponds to the set of vertices of a graph. To link the vertices, we consider the transitions between states $a^i_{mm'}$ as the graph edges. As a result, truncating the number of eigenvectors to $L = 2$ or $L = 3$ results in a two- or three-dimensional graph depicting the stochastic dynamic of a time-series.

5.2 Results

For the sake of visualizing the separability between ADHD and control subjects, we applied Low-dimensional HMM visualization on the kernel matrix resulting. Figure 5.1 illustrates the data projected over the two-channel configurations with the optimal parameters at the best performing block for each condition. As shown in the figure using the whole set of channels, ADHD is more concentrated in one region, and between the two classes, no clear boundary is identified, while COI data is more dispersed, which proves that classes are more cluttered in COI than in all channels so that similarity kernel better defines a boundary between both classes in the latter condition.

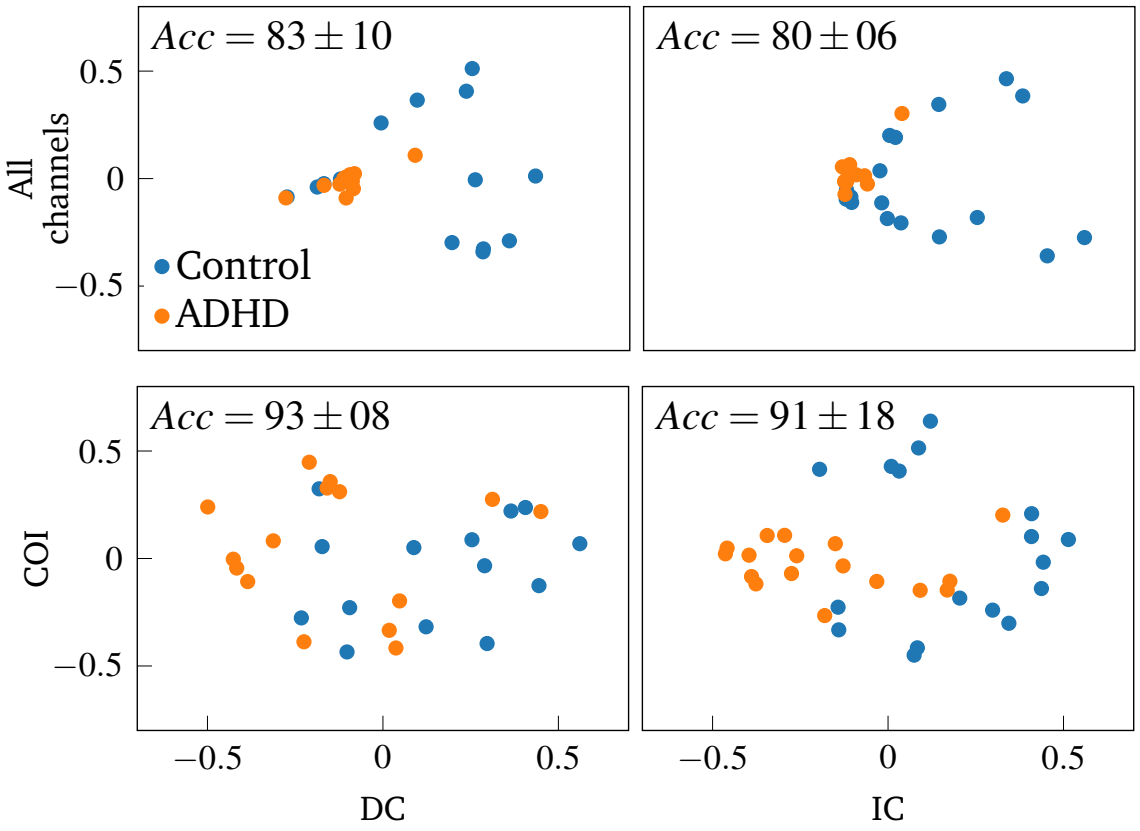


Figure 5.1. 2D representations for the HMMs in eigth COI and all channels in both conditions.

Similarly, we performed the visualization of low-dimensional state for control and ADHD subjects in *Theta*, *Alpha*, and *Beta* frequency bands as shown in Figure 5.2. The graphs reflect that the best performances are in an increasing condition where the *Theta* and *Beta* bands show hits above 90%, which is consistent with the scientific evidence of these frequency spectra for the diagnosis of ADHD. On the other hand, although in DC the data distributions are more dispersed and separability between classes, in IC the data are concentrated and delimited, favoring the classification process.

For the sake of methodology interpretability, we apply the proposed HMM visualization in Section 5.1.3 to the HMMs representing each subject in the datasets, resulting in a two-dimensional location of each state. Figures 5.3 to 5.5 illustrate the resulting HMM representations for the 32 channels montage, the COI, and the *Beta* rhythm with COI, respectively. In either case, we fixed the number of states and ρ power at their optimal value. The analysis excluded other reward conditions, blocks, and rhythms as the low performance restrain the interpretability. Since each state models the EEG channels as a multivariate Gaussian, we represent the states as head topographic maps plotting the mean EEG activity $\boldsymbol{\mu}_{im}$, which allows a straightforward interpretation of signal dynamics. Edges

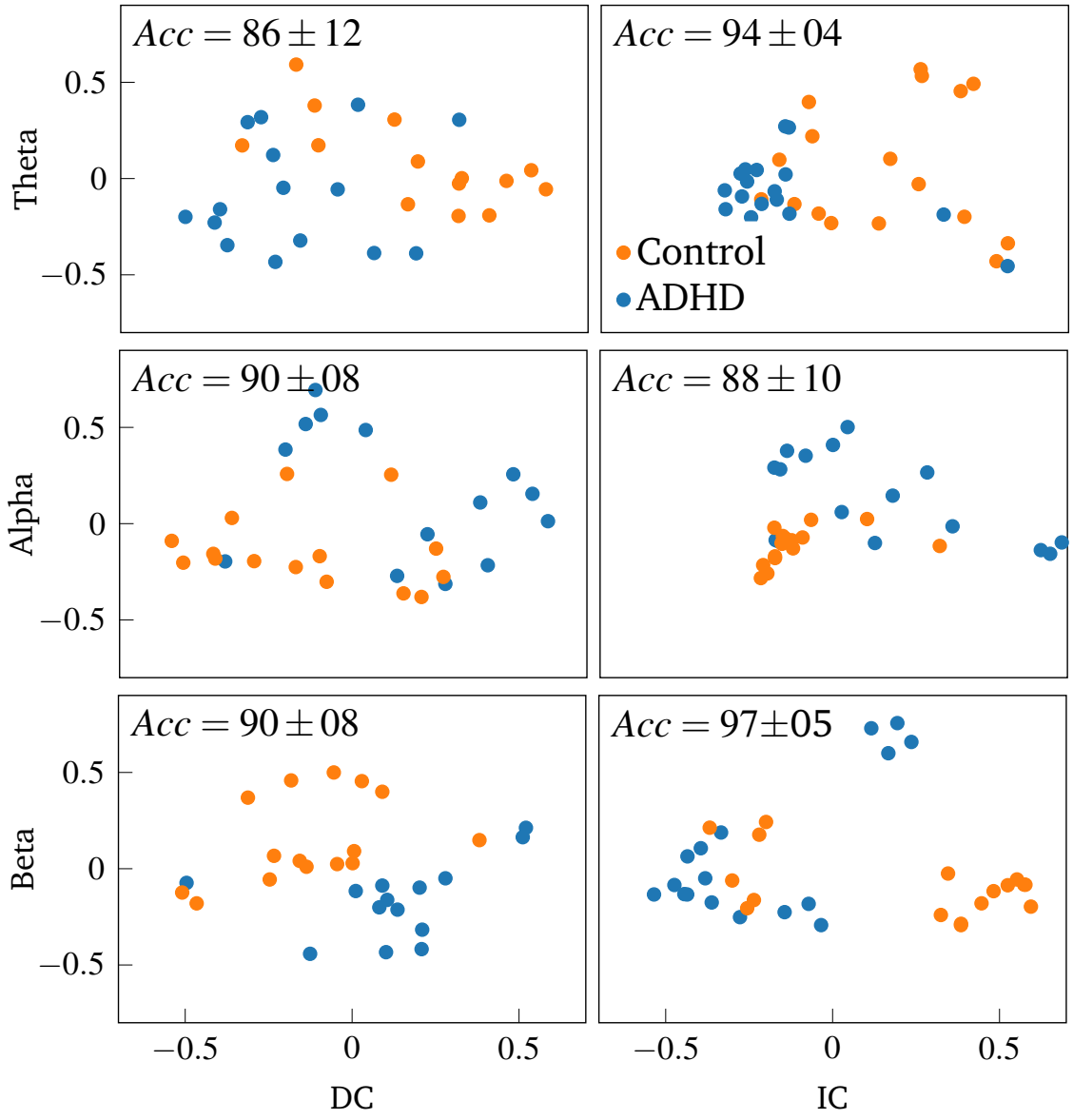
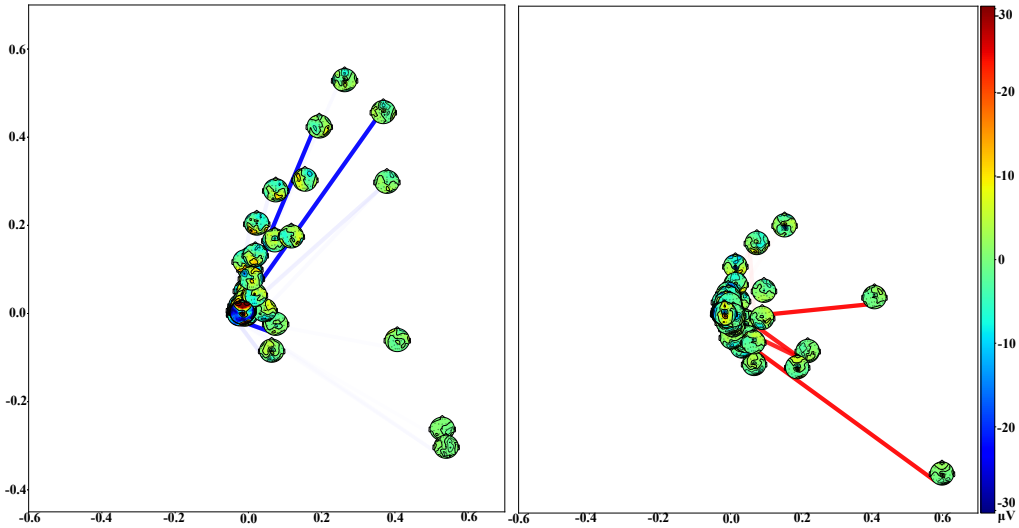


Figure 5.2. 2D representation for HMMs in frequency rhythms with eight channels in both conditions.



(a) Healthy control children (b) ADHD children

Figure 5.3. 2D HMM representation for the 32 channel montage and the decreasing condition. Colorbar denotes the EEG amplitudes (μV) for the topographic maps.

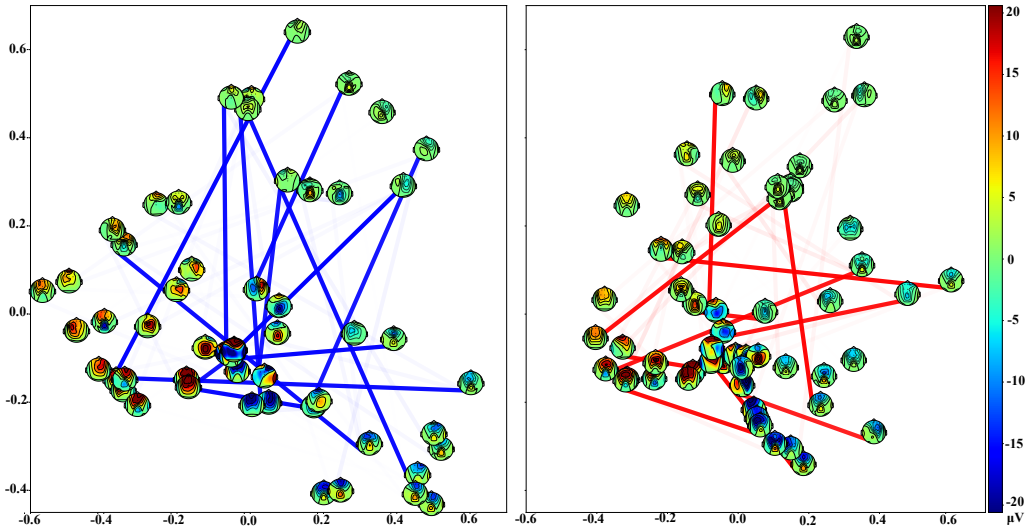
linking the states result from the transition matrices representing the EEG dynamics so that the opacity is directly proportional to the transition probability, and less likely transitions are less visible. The red and blue links denote ADHD and HC state transitions, respectively.

The resulting HMM graphs for 32 channels montage in Figure 5.3 presents most of the states concentrated in a region with a few dynamic changes evidenced through the transitions. For the COI, Figure 5.4 exhibits more changes in brain electrical activity. Nonetheless, transition paths lack any discriminative pattern and behave randomly. In contrast, Figure 5.5 evidences two benefits of the *Beta* band. Firstly, states widely locate over the 2D space to gather similar EEG potentials and separate different ones. Secondly, links allow discrimination since ADHD transitions concentrate over the plot center, whereas HC ones surround it. Therefore, the spatial channel selection and spectral filtering change the EEG data distribution and highlight its dynamics to enhance class separability.

5.3 Discussion

To develop a methodology for visualizing stochastic representations that facilitates the interpretability of inference machines, in this chapter, we present a methodology that allows interpreting the coding of HMMs trained with EEG recordings using dimension reduction.

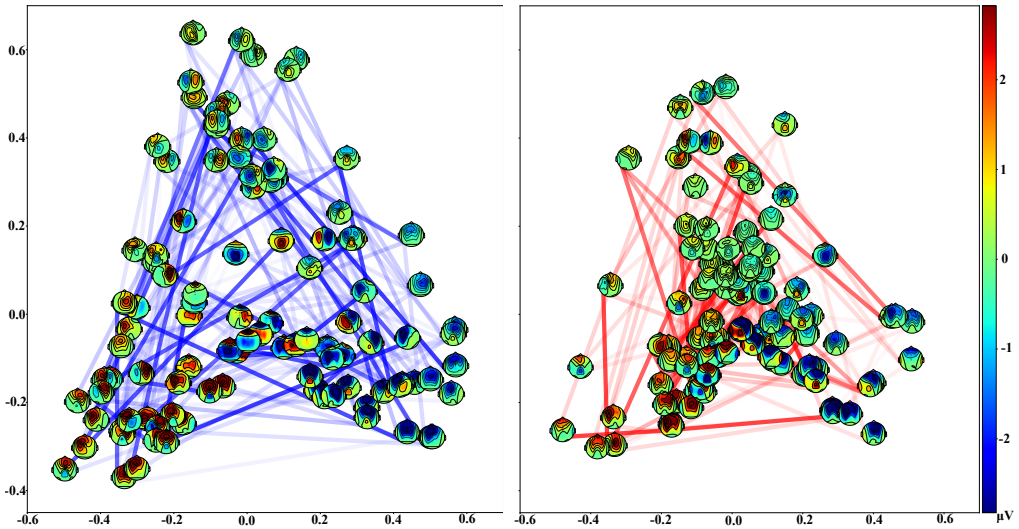
To reach a spatial interpretation of the HMMs, The Figures 5.1 and 5.3 supports the hypothesis that by feeding the model with all the available channels of a given EEG dataset,



(a) Healthy control children (b) ADHD children

Figure 5.4. 2D HMM representation for the eight COI montage and the decreasing condition. Colorbar denotes the EEG amplitudes (μV) for the topographic maps.

there is no differences allowing to distinguish between two classes dynamic changes since the observed distribution over each state are localized in the same zone. Conversely, in the Figure 5.4 when working on selected channels, results draw more dispersed regions with similar electrophysiological activity although with undefined transitions. However, Figure 5.2 shows a greater separability between the classes when filtering EEG recordings in frequency bands, highlighting that the band in the *Beta* band, clearer transitions and regions are identified between control and ADHD subjects as evidenced Figure 5.5. In the latter group, predominantly transitions of neutral fronto-central activity are concentrated in the center of the graph. On the contrary, control subjects exhibited transitions condensed in the periphery, with a greater probability of transfer to lateral polarizations from positive and negative fronto-central activities.



(a) Healthy control children (b) ADHD children

Figure 5.5. 2D HMM representation for the eight COI montage, *Beta* rhythm and the increasing condition. Colorbar denotes the EEG amplitudes (μV) for the topographic maps.

Chapter 6

Conclusions

This research proposes a supported diagnostic methodology for ADHD, based on HMM as a stochastic model. According to the RSST paradigm, we compare the results of behavioral data classification with the similarity between EEG records in two-channel configurations for each reward and condition block, followed by spectral analysis in *Theta*, *Alpha*, and *Beta* bands. Thanks to the proposed methodology, we evaluate the similarity between EEG signals from their HMMs using closed-form PPK computation for a sequence length and a number of hidden states, and fast PKK computation for each sequence length, power *rho*, and a number of hidden states. Thus, given the inner product characteristics, the resulting similarity kernel can feed a classifier such as a support vector machine. Finally, with the motivation to spatially interpret the HMM, we projected the similarity kernel into a low-dimensional space to visualize the separability between ADHD and control subjects. Also, we perform a similarity kernel between the probability density functions of each HMM state that feeds a KPCA, where each point is represented by a topo-plot with the means of the Gaussian distribution of each EEG, and the changes between such electrical activity are shown through lines that reflect how likely is the transition between these states.

According to the results, HMMs are suitable tools for encoding multidimensional features of the EEG signal, and together with PPK results in a sufficiently complex measure that improves classification performance. We identified that a Bhattacharya Kernel does not achieve the highest accuracy. Thus, finding the optimal parameters by fast PPK computation allows for greater separability between classes by favoring less likely events that are more discriminative in the classification task.

We identified that behavioral data are not sufficient as a diagnostic support tool evidence that our methodology exceeds the results with behavioral data, which agrees that neuropsychological tests do not allow correct discrimination of ADHD [70].[71] Our results support the need to incorporate electrophysiological biomarkers to accurately distinguish ADHD from control children. Also, we identify that the best performance in the group of selected channels is in line with functional neuroanatomic findings, already described in the literature about frontal and prefrontal circuitry disorders in children with ADHD. Our

research replicates and supports these findings.

Consequently, this finding supports the already stated hypothesis that inhibitory performance in ADHD children is highly sensitive to the effect of reward. Our results highlight a crucial role of motivation, that goes beyond a plain poor impulse control, contributing to a better understanding of the underpinnings mechanisms of this condition. Also, this finding supports the role of *Theta* and *Beta* bands as consistent hallmarks of ADHD biomarkers[75].[76] The lack of differences over *Alpha* band exist less difference between classes is also consistent with what has already been described for ADHD children. *Beta* band obtained the best classification performance through the four blocks, which supports the assumption that this frequency rhythm may be correlated to differences between ADHD and controls in regard to attentional allocation during the execution of cognitive tasks. Likewise, inhibitory paradigms, as the RSST, allows to gather information about an event related potential such as the ERN, which features may change in front of motivational contingencies such as the presence of reward and punishment .[74]

Within the implications of this methodology, the specialist will project the subject's HMM to a point in the low-dimensional representation space, identifying the patient's proximity to the bounded regions of both classes. Also, it will facilitate therapeutic or diagnostic decisions, which will allow a longitudinal analysis of the patient.

Bibliography

[1] S. Khoshnoud, M. A. Nazari, and M. Shamsi, "Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals," *Journal of integrative neuroscience*, 2018.

[2] V. del Barrio, "Diagnostic and Statistical Manual of Mental Disorders," in *Encyclopedia of Applied Psychology, Three-Volume Set*, 2004.

[3] S. M. Snyder, T. A. Rugino, M. Hornig, and M. A. Stein, "Integration of an EEG biomarker with a clinician ' s ADHD evaluation," vol. 330, pp. 1–17, 2015.

[4] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD," *Clinical EEG and Neuroscience*, vol. 41, no. 1, pp. 1–10, 2010.

[5] L. Marquardt, H. Eichele, A. J. Lundervold, J. Haavik, and T. Eichele, "Event-related-potential (ERP) correlates of performance monitoring in adults with attention-deficit hyperactivity disorder (ADHD)," *Frontiers in psychology*, vol. 9, p. 485, 2018.

[6] J. F. Saad, M. R. Kohn, S. Clarke, J. Lagopoulos, and D. F. Hermens, "Is the Theta/Beta EEG Marker for ADHD Inherently Flawed?" *Journal of Attention Disorders*, vol. 22, no. 9, pp. 815–826, 2018.

[7] M. J. Groom, G. Scerif, P. F. Liddle, M. J. Batty, E. B. Liddle, K. L. Roberts, J. D. Cahill, M. Liotti, and C. Hollis, "Effects of motivation and medication on electrophysiological markers of response inhibition in children with attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 67, no. 7, pp. 624–631, 2010.

[8] S. Whitmont, R. Meares, E. Gordon, I. Lazzaro, and S. Clarke, "The Modulation of Late Component Event Related Potentials by Pre-Stimulus EEG Theta Activity in ADHD," *International Journal of Neuroscience*, vol. 107, no. 3-4, pp. 247–264, 2008.

[9] J. Tornero Lucas, "Machine Learning: Modelos Ocultos de Markov (HMM) y Redes Neuronales Artificiales (ANN)," p. 53, 2017. [Online]. Available: http://diposit.ub.edu/dspace/bitstream/2445/122446/2/memoria.pdf

[10] B.-H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden markov models," *AT&T technical journal*, vol. 64, no. 2, pp. 391–408, 1985.

[11] C. Lu, J. M. Schwier, R. M. Craven, L. Yu, R. R. Brooks, and C. Griffin, "A normalized statistical metric space for hidden Markov models," in *IEEE Transactions on Cybernetics*, 2013.

[12] M. Falkhausen, H. Reininger, and D. Wolf, "Calculation of distance measures between hidden Markov models." *Simulation*, 1995.

[13] J. Zeng, J. Duan, and C. Wu, "A new distance measure for hidden Markov models," *Expert Systems with Applications*, 2010.

[14] E. Epaillard and N. Bouguila, "Data-free metrics for dirichlet and generalized dirichlet mixture-based hmms – a practical study," *Pattern Recognition*, vol. 85, pp. 207 – 219, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031830311X

[15] P. Ruiz, C. Philbrick, and P. Sauer, "Modeling approaches for computational cost reduction in stochastic unit commitment formulations," *Power Systems, IEEE Transactions on*, vol. 25, pp. 588–589, 02 2010.

[16] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of BMD and ADHD patients using their EEG signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956–1963, 2011.

[17] B. Abibullaev and J. An, "Decision support algorithm for diagnosis of ADHD using electroencephalograms," *Journal of medical systems*, vol. 36, no. 4, pp. 2675–2688, 2012.

[18] A. Lenartowicz and S. K. Loo, "Use of EEG to Diagnose ADHD," 2014.

[19] S. Itani, M. Rossignol, F. Lecron, and P. Fortemps, "Towards interpretable machine learning models for diagnosis aid: a case study on attention deficit/hyperactivity disorder," *PloS one*, vol. 14, no. 4, p. e0215720, 2019.

[20] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[21] E. Topol, *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[22] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[23] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 559–560.

[24] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017.

[25] M. Fredriksen, A. A. Dahl, E. W. Martinsen, O. Klungsoyr, S. V. Faraone, and D. E. Peleikis, "Childhood and persistent ADHD symptoms associated with educational failure and long-term occupational disability in adult ADHD." *Attention deficit and hyperactivity disorders*, vol. 6, no. 2, pp. 87–99, jun 2014.

[26] J. Biederman, S. V. Faraone, T. J. Spencer, E. Mick, M. C. Monuteaux, and M. Aleardi, "Functional impairments in adults with self-reports of diagnosed ADHD: A controlled study of 1001 adults in the community." *The Journal of clinical psychiatry*, vol. 67, no. 4, pp. 524–540, apr 2006.

[27] M. Brod, B. Pohlman, R. Lasser, and P. Hodgkins, "Comparison of the burden of illness for adults with ADHD across seven countries: a qualitative study." *Health and quality of life outcomes*, vol. 10, p. 47, may 2012.

[28] A. Cussen, E. Sciberras, O. C. Ukoumunne, and D. Efron, "Relationship between symptoms of attention-deficit/hyperactivity disorder and family functioning: a community-based study." *European journal of pediatrics*, vol. 171, no. 2, pp. 271–280, feb 2012.

[29] K. L. Humphreys and S. S. Lee, "Risk Taking and Sensitivity to Punishment in Children with ADHD, ODD, ADHD+ODD, and Controls." *Journal of psychopathology and behavioral assessment*, vol. 33, no. 3, pp. 299–307, 2011.

[30] C. Gómez-Restrepo, C. Escudero, D. Matallana, L. González, and V. Rodriguez, "Encuesta Nacional de Salud Mental," *MInsalud. Colciencias*, vol. 1, 2015.

[31] C. Hidalgo-López, A. M. Gómez-Álzate, J. García-Valencia, and J. D. Palacio-Ortiz, "Riesgo de trastorno por déficit de atención e hiperactividad y otros trastornos psiquiátricos de los hermanos de pacientes con TDAH," *Revista Colombiana de Psiquiatría*, vol. 48, no. 1, pp. 44–49, 2019.

[32] S. Mnookin, "Out of The Shadows: Making Mental Health a Global Development Priority," *Ecosystems and Human Well-being: A Framework for Assessment*, 2016.

[33] Y. T. De Galvis, "Costos asociados con la salud mental," 2018.

[34] CONSEJO NACIONAL DE POLÍTICA ECONÓMICA Y SOCIAL CONPES, "ESTRATEGIA PARA LA PROMOCIÓN DE LA SALUD MENTAL EN COLOMBIA," Bogotá,Colombia, p. 70, 2020. [Online]. Available: https://colaboracion.dnp.gov.co/CDT/Conpes/Econ{\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{o\global\mathchardef\accent@spacefactor\spacefactor}\let\begingroup\endgroup\relax\let\ignorespaces\relax\accent19o\egroup\spacefactor\accent@spacefactor}micos/3992.pdf

[35] N. T. Martínez, C. J. R. Rodríguez, C. de Santacruz, N. B. Bautista, J. Collazos, and C. Gómez–Restrepo, "Problemas mentales, trastornos del afecto y de ansiedad en la población desplazada por la violencia en Colombia, resultados de la Encuesta Nacional de Salud Mental 2015," *Revista colombiana de psiquiatría*, vol. 45, pp. 113–118, 2016.

[36] J. M. Berg, R. D. Latzman, N. G. Bliwise, and S. O. Lilienfeld, "Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the UPPS for psychopathology," *Psychological assessment*, 2015.

[37] F. Di Michele, L. Prichep, E. R. John, and R. J. Chabot, "The neurophysiology of attention-deficit/hyperactivity disorder," *International Journal of Psychophysiology*, vol. 58, no. 1, pp. 81–93, 2005.

[38] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (5th Edition)*, 2013.

[39] D. W. Dunn and W. G. Kronenberger, "Attention-deficit / hyperactivity disorder in children and adolescents," vol. 21, pp. 933–940, 2003.

[40] F. H. Duffy, A. Shankardass, G. B. Mcanulty, and H. Als, "A unique pattern of cortical connectivity characterizes patients with attention deficit disorders : a large electroencephalographic coherence study," pp. 1–19, 2017.

[41] T. W. Picton, "The P300 Wave of the Human Event-Related Potential," no. November 1992, 2017.

[42] W. Hospital and W. Hospital, "THE MODULATION OF LATE COMPONENT EVENT RELATED POTENTIALS BY PRE-STIMULUS EEG THETA ACTIVITY IN ADHD," vol. 107, 2001.

[43] S. V. Faraone and C. Bonvicini, "Biomarkers in the Diagnosis of ADHD – Promising Directions," 2014.

[44] J. Picardo and G. García, "Hypoarousal non-stationary ADHD biomarker based on echo- state networks," no. 2017, 2018.

[45] R. J. Chabot and G. Serfontein, "Quantitative Electroencephalographic Profiles of Children with Attention Deficit Disorder," vol. 3223, no. 95, 1996.

[46] D. A. Pineda, F. Lopera, J. D. Palacio, D. Ramirez, and G. C. Henao, "Prevalence estimations of attention-deficit/hyperactivity disorder: Differential diagnoses and comorbidities in a Colombian sample," *International Journal of Neuroscience*, 2003.

[47] C. Sridhar, S. Bhat, U. R. Acharya, H. Adeli, and G. M. Bairy, "Diagnosis of attention deficit hyperactivity disorder using imaging and signal processing techniques," 2017.

[48] R. J. Barry, A. R. Clarke, and S. J. Johnstone, "A review of electrophysiology in attention-deficit / hyperactivity disorder : I . Qualitative and quantitative electroencephalography," vol. 114, pp. 171–183, 2003.

[49] I. I. Event-related, R. J. Barry, S. J. Johnstone, and A. R. Clarke, "A review of electrophysiology in attention-deficit / hyperactivity disorder :," vol. 114, pp. 184–198, 2003.

[50] J. R. Wiersema, J. J. V. D. Meere, and H. Roeyers, "ERP correlates of impaired error monitoring in children with ADHD," pp. 1417–1430, 2005.

[51] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data," *IEEE transactions on cybernetics*, vol. 45, no. 12, pp. 2668–2679, 2015.

[52] A. Mueller, G. Candrian, V. A. Grane, J. D. Kropotov, V. A. Ponomarev, and G.-M. Baschera, "Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study," *Nonlinear biomedical physics*, vol. 5, no. 1, pp. 1–18, 2011.

[53] D. Fair, J. T. Nigg, S. Iyer, D. Bathula, K. L. Mills, N. U. F. Dosenbach, B. L. Schlaggar, M. Mennes, D. Gutman, and S. Bangaru, "Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data," *Frontiers in systems neuroscience*, vol. 6, p. 80, 2013.

[54] A. Allahverdy, A. M. Nasrabadi, and M. R. Mohammadi, "Detecting ADHD children using symbolic dynamic of nonlinear features of EEG," in *2011 19th Iranian Conference on Electrical Engineering*. IEEE, 2011, pp. 1–4.

[55] S. Furlong, J. R. Cohen, J. Hopfinger, J. Snyder, M. M. Robertson, and M. A. Sheridan, "Resting-state eeg connectivity in young children with adhd," *Journal of Clinical Child & Adolescent Psychology*, pp. 1–17, 2020.

[56] S. Galindo-Noreña, D. Cárdenas-Peña, and A. A. Orozco-Gutierrez, "Csp-based discriminative capacity index from eeg supporting adhd diagnosis," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1343–1347.

[57] J. L. Marcano, M. A. Bell, and A. A. L. Beex, "Classification of ADHD and Non-ADHD Subjects Using a Universal Background Model," *Biomedical signal processing and control*, vol. 39, pp. 204–212, jan 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31186670https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6557459/

[58] S. Solhjoo, A. M. Nasrabadi, and M. R. H. Golpayegani, "Classification of chaotic signals using hmm classifiers: Eeg-based mental task classification," in *2005 13th European Signal Processing Conference*. IEEE, 2005, pp. 1–4.

[59] C. Stam, B. Jelles, H. Achtereekte, S. Rombouts, J. Slaets, and R. Keunen, "Investigation of eeg non-linearity in dementia and parkinson's disease," *Electroencephalography and clinical neurophysiology*, vol. 95, no. 5, pp. 309–317, 1995.

[60] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *Journal of Machine Learning Research*, 2004.

[61] T. Jebara and R. Kondor, "Bhattacharyya and expected likelihood kernels," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2003.

[62] C. S. van Meel, D. J. Heslenfeld, J. Oosterlaan, M. Luman, and J. A. Sergeant, "Erps associated with monitoring and evaluation of monetary reward and punishment in children with adhd," *Journal of Child Psychology and Psychiatry*, vol. 52, no. 9, pp. 942–953, 2011.

[63] W. J. Gehring, Y. Liu, J. M. Orr, and J. Carp, "The error-related negativity (ern/ne)." 2012.

[64] G. Hajcak, J. S. Moser, N. Yeung, and R. F. Simons, "On the ern and the significance of errors," *Psychophysiology*, vol. 42, no. 2, pp. 151–160, 2005.

[65] S. Wang, Y. Yang, W. Xing, J. Chen, C. Liu, and X. Luo, "Altered neural circuits related to sustained attention and executive control in children with adhd: an event-related fmri study," *Clinical Neurophysiology*, vol. 124, no. 11, pp. 2181–2190, 2013.

[66] F. X. Castellanos and E. Proal, "Large-scale brain systems in adhd: beyond the prefrontal–striatal model," *Trends in cognitive sciences*, vol. 16, no. 1, pp. 17–26, 2012.

[67] S. Watanabe, A. Nakamura, and B. H. Juang, "Structural bayesian linear regression for hidden markov models," *Journal of Signal Processing Systems*, vol. 74, no. 3, pp. 341–358, 2014.

[68] A. C. Kelly, A. Scheres, E. S. SONUGA-BARKE, and F. X. CASTELLANOS, "11 functional neuroimaging of reward and motivational pathways in adhd," *Handbook of attention deficit hyperactivity disorder*, p. 209, 2007.

[69] L. Chen and H. Man, "Fast schemes for computing similarities between gaussian hmms and their applications in texture image classification," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 13, p. 164742, 2005.

[70] J. Holmes, S. E. Gathercole, M. Place, T. P. Alloway, J. G. Elliott, and K. A. Hilton, "The diagnostic utility of executive function assessments in the identification of adhd in children," *Child and Adolescent Mental Health*, vol. 15, no. 1, pp. 37–43, 2010.

[71] R. A. Barkley, "Neuropsychological testing is not useful in the diagnosis of adhd: Stop it (or prove it)!" *The ADHD Report*, vol. 27, no. 2, pp. 1–8, 2019.

[72] N. D. Volkow, G.-J. Wang, J. H. Newcorn, S. H. Kollins, T. L. Wigal, F. Telang, J. S. Fowler, R. Z. Goldstein, N. Klein, J. Logan *et al.*, "Motivation deficit in adhd is associated with dysfunction of the dopamine reward pathway," *Molecular psychiatry*, vol. 16, no. 11, pp. 1147–1154, 2011.

[73] E. B. Liddle, C. Hollis, M. J. Batty, M. J. Groom, J. J. Totman, M. Liotti, G. Scerif, and P. F. Liddle, "Task-related default mode network modulation and inhibitory control in adhd: Effects of motivation and methylphenidate," *Journal of Child Psychology and Psychiatry*, vol. 52, no. 7, pp. 761–771, 2011.

[74] I. Marx, T. Hacker, X. Yu, S. Cortese, and E. Sonuga-Barke, "Adhd and the choice of small immediate over larger delayed rewards: a comparative meta-analysis of performance on simple choice-delay and temporal discounting paradigms," *Journal of attention disorders*, p. 1087054718772138, 2018.

[75] M. Arns, C. K. Conners, and H. C. Kraemer, "A decade of eeg theta/beta ratio research in adhd: a meta-analysis," *Journal of attention disorders*, vol. 17, no. 5, pp. 374–383, 2013.

[76] J. Van Doren, H. Heinrich, M. Bezold, N. Reuter, O. Kratz, S. Horndasch, M. Berking, T. Ros, H. Gevensleben, G. H. Moll *et al.*, "Theta/beta neurofeedback in children with adhd: feasibility of a short-term setting and plasticity effects," *International Journal of Psychophysiology*, vol. 112, pp. 80–88, 2017.